МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
"Харківський політехнічний інститут"

О. Б. Білоцерківський, Н. В. Ширяєва

ЕКОНОМЕТРІЯ

Затверджено
редакційно-видавничою
радою університету,
протокол № 2 від 27.06.08

Харків НТУ „ХПІ” 2008

ББК 22.213
Б 78
УДК 534.1

Рецензенти: *Г. Є. Мазнєв*, канд. екон. наук, проф. ХНТУСГ;
О. М. Гаврись, канд. екон. наук, проф. НТУ "ХПІ"

Б 78 **Білоцерківський О. Б.** Економетрія: навч.-метод. посібник / О. Б. Білоцерківський, Н. В. Ширяєва. – Харків: НТУ "ХПІ", 2008. – 80 с.

ISBN

Посібник містить основи економетрії, включаючи методи визначення параметрів лінійних рівнянь парної і множинної регресії, нелінійних рівнянь, часових рядів, оцінки їх тісноти та значущості. Наводяться приклади вирішення розрахункових завдань і варіанти для самостійної роботи студентів.

Призначений для студентів спеціальності 7.050206 "Менеджмент зовнішньоекономічної діяльності", а також буде корисним для студентів інших економічних спеціальностей.

Іл. 3. Табл. 24. Бібліогр. 7 назв.

ББК 22.213

ISBN

ЗМІСТ

ПЕРЕДМОВА

Особливістю нинішнього етапу розвитку вітчизняної економіки в умовах переходу до ринкових відносин є зростання інтересу фахівців до наукового вирішення проблем з використанням економіко-математичних методів і побудованих на їхній основі моделей. Це проявляється, насамперед, у тому, що математичні методи й моделі в економіці вимагають ретельного урахування всіх можливих ситуацій, що робить управлінські рішення науково обґрунтованими для забезпечення збалансованого та стійкого функціонування господарського механізму. Застосування сучасних методів дослідження економічних процесів і явищ дозволяє повніше та глибше обґрунтувати темпи й пропорції розвитку на макро- та мікрорівні, домогтися оптимуму серед альтернативних рішень. При цьому зростає роль економетрії як науки про виміри в економіці й керуванні з використанням сучасних економіко-математичних методів, моделей і засобів їхньої реалізації.

У даному посібнику викладаються теоретичні основи курсу «Економетрія». Велика увага приділена парній і множинній регресії, класичному методу найменших квадратів і аналізу часових рядів. Розглядаються різні аспекти багатомірної регресії: мультиколінеарність, фіктивні змінні, лінеаризація моделі, часткова кореляція.

Посібник містить приклади вирішених завдань. Для закріплення знань з курсу та придбання навичок, необхідних для побудови й аналізу економетричних моделей, для студентів очної форми навчання передбачені розрахунково-графічні завдання за основними темами курсу. Номер варіанта збігається з порядковим номером прізвища студента за списком у журналі групи й реєструється викладачем при видачі завдань.

1. ПРЕДМЕТ, МЕТОДИ ТА ЗАВДАННЯ ЕКОНОМЕТРІЇ. ОСОБЛИВОСТІ ПОБУДОВИ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ

Економетрія – це один з базових курсів підготовки економістів. Ця дисципліна заснована на фундаменті знань економічної теорії, матричної алгебри, теорії математичної й загальної статистики.

1.1. Предмет, методи та завдання економетрії

Економетрія – це один з напрямків економіко-математичних методів аналізу, що полягає в статистичному вимірі (оцінюванні) параметрів математичних виразів, які характеризують деяку економічну концепцію про взаємозв'язок і розвиток об'єкта, явища, що необхідно для одержання конкретних економічних висновків на основі економетричних моделей.

Економетрика – це наука, що досліджує кількісні закономірності й взаємозалежності в економіці за допомогою методів математичної статистики. Основа цих методів – кореляційно-регресійний аналіз. Використання сучасних методів математичної статистики почалося в біології. В останній чверті XIX століття англійський біолог К. Пірсон поклав початок сучасній математичній статистиці вивченням кривих розподілу числових характеристик людського організму. Потім він та його школа перейшли до вивчення кореляцій у біології й побудові лінійних функцій регресії. Перші роботи з економетрики з'явилися наприкінці XIX – початку XX століття. В 1897 р. з'явилася робота одного із засновників математичної школи в економічній теорії В. Парето, присвячена статистичному вивченню доходів населення в різних країнах. Була запропонована крива Парето $y = A(x-a)^{\alpha}$, де x – величина доходу; y – кількість осіб, що мають дохід, більший x; a – мінімальний дохід; A та α – параметри залежності, одержувані статистичними методами.

На початку XX століття вийшло кілька робіт англійського статистика Гукера, у яких він застосував кореляційно-регресійні методи, що розроблені Пірсоном та його школою, для вивчення взаємозв'язків економічних показників, зокрема – впливу числа банкрутств на товарній біржі на ціну зерна. У роботах Гукера проводилася ідея часового лага між економічними змінними, а також ідея кореляційного аналізу не самих величин, а їхніх збільшень. Надалі з'явилася величезна кількість робіт як щодо розвитку теорії математичної статистики та її прикладних елементів, так і щодо практичного використання цих методів в економічному аналізі. До першої групи можуть бути віднесені, наприклад, роботи Р. Фішера по дисперсійному аналізу, до другої групи – роботи з оцінки й дослідження виробничих функцій, зокрема – класична робота Кобба й Дугласа (1928 р.).

Розглянемо ряд відомих економічних законів і положень, в основі яких лежать економетричні дослідження.

Крива Філіпса – це крива, що відбиває взаємозв'язок між рівнем інфляції й рівнем безробіття. Олбан Філіпс (Phillips) (1914–1975 рр.) – австралійський економіст, що працював в Англії.

У 1958 р. О. Філіпс, спираючись на емпіричні дані по Англії за 1861–1957 рр., вивів кореляційну залежність між рівнем безробіття й зміною приросту грошової заробітної плати. Ця залежність була представлена у вигляді кривої, що характеризує функціональний зв'язок цих двох величин: чим вище безробіття, тим менше приріст грошової заробітної плати, тим нижче ріст цін; і навпаки, чим нижче безробіття й вище зайнятість, тим більше приріст грошової заробітної плати, тим вище темп росту цін. «Криві Філіпса» були покладені в основу кейнсіанської теорії інфляції, у якій зв'язок між рівнем зайнятості й інфляцією встановлювався через динаміку грошової заробітної плати. У загальному вигляді залежність, відбиту кривою Філіпса, можна записати як

$$\pi_t - \pi_{t-1} = -\beta \cdot \left(u_t - u_t^e\right) + \varepsilon_t,$$

де π_t – рівень інфляції в момент часу t; u_t – рівень безробіття в момент часу t; u_t^e – рівень природного безробіття; ε_t – вплив шокових змін; β – коефіцієнт або параметр моделі.

Графік залежності, вираженої кривою Філіпса, представлений на рис. 1.1.

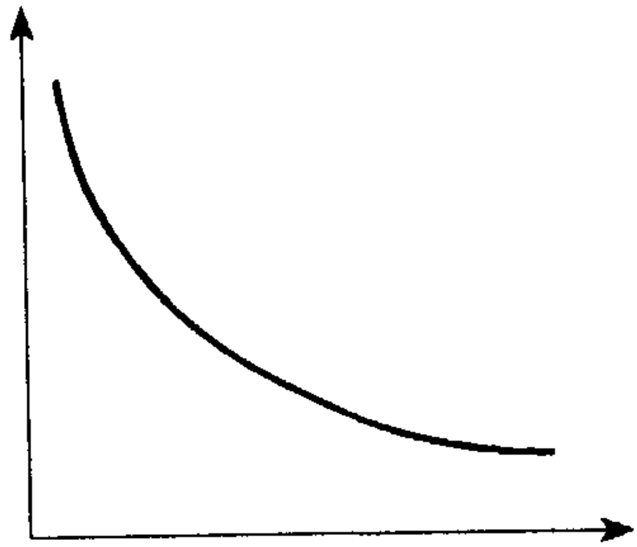

Рисунок 1.1 – Крива Філіпса

Можна помітити, що не завжди збільшення рівня безробіття приводить до зниження інфляції, – для ряду країн визначаються такі періоди, коли обидва ці показники збільшуються одночасно. Тому крива Філіпса вивчалася для різних економічних систем й у різні періоди часу. Наприклад, ця крива досить точно відбиває історію інфляції та безробіття у США 60-х років. Однак аналіз статистичних даних першої половини 70-х років свідчить про те, що криву Філіпса не завжди можна застосовувати. У 1973–1975 рр. в США спостерігалося одночасне збільшення інфляції й безробіття. В 1976–1979 рр. знову економіка розвивалась згідно з кривою Філіпса.

Крива Лаффера. Артур Лаффер (Laffer) – сучасний американський економіст. Залежність, відбита кривою Лаффера, була відкрита у 80-х роках XX сторіччя в процесі дослідження функцій, які б характеризували залежність бюджетних надходжень від ставок податку на прибуток і заробітну плату. Було доведено, що існує довгострокова залежність між ставками податків і надходженнями до бюджету, та, крім того, існує оптимальний рівень оподатковування, при якому функція досягає свого максимуму. Були здійснені економетричні дослідження за допомогою кривої Лаффера для різних країн.

Крива Лаффера показує нелінійний зв'язок між податковою ставкою (в процентах) та надходженнями від податків у бюджет (рис. 1.2). Формально дана залежність може бути виражена у вигляді

$$y = a+bx+cx^2,$$

де $c < 0$; x – податкова ставка; y – надходження від податків до бюджету.

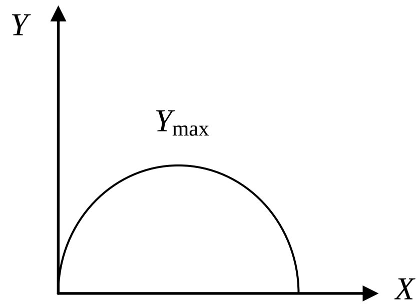

Рисунок 1.2 – Крива Лаффера

Наведені приклади свідчать про досить широке застосування економетричних методів для обґрунтування економічних законів і положень. Саме синтез математики, економіки й статистики зробив економетрію потужним інструментом для теоретичного й прикладного аналізу різних соціально-економічних процесів.

Методологія економетричного моделювання тісно пов'язана з використанням системного підходу. Соціально-економічна система може бути представлена нескінченним числом структурних і функціональних інваріантів, що відбивають взаємозв'язки між різними процесами, які протікають у цій системі (економічними, соціальними, екологічними, демографічними та ін.). Система описується за допомогою якісних і кількісних характеристик, названих параметрами. Параметри становлять основу мов опису систем, а при формалізації ототожнюються з незалежними змінними математичного опису процесу функціонування систем.

При побудові економетричної моделі реалізується метод моделювання за принципом «чорного ящика», коли дослідникові невідомий механізм процесів, що протікають у системі, вивчити який можна за вхідними і вихідними характеристиками системи. Вхідні та вихідні характеристики системи часто ототожнюють із екзогенними й ендогенними змінними, або в кореляційно-регресійному аналізі вживають терміни «незалежні (факторні) змінні або ознаки» й «залежні (результативні) змінні або ознаки». Графічно принцип «чорного ящика» зображений на рис. 1.3.

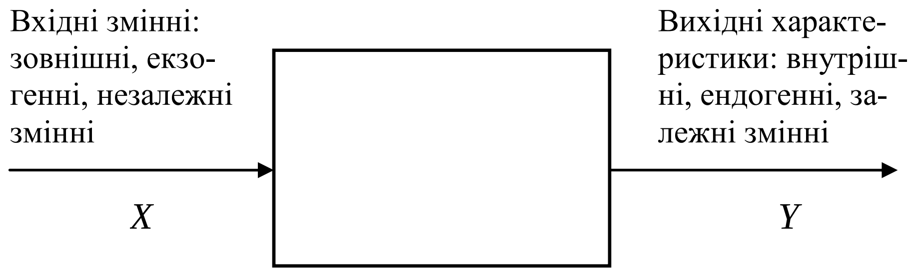

Рисунок 1.3 – Дослідження системи за принципом «чорного ящика»

Дослідникові необхідно виділити вхідні й вихідні характеристики та на підставі економетричних методів установити характер причинно-наслідкових зв'язків, що лежать в основі механізму функціонування соціально-економічної системи.

Заключним етапом у процедурі послідовної формалізації опису процесів функціонування соціально-економічних систем є розробка математичних моделей.

Математична модель – це вираження формальної залежності у вигляді деяких математичних співвідношень (функцій, систем рівнянь або нерівностей, операторів та ін.), що відбивають зв'язок між певними явищами. Серед класів економіко-математичних моделей, що відбивають функціонування соціально-економічної системи, необхідно виділити економетричні моделі.

1.2. Особливості побудови економетричної моделі

Економетрична модель – це особливий клас економіко-математичних моделей, у яких дослідник вирішує цілий ряд завдань:

- вибір форми математичної залежності, що описує поводження економічного об'єкта на основі системи спостережень;
- оцінка параметрів даної моделі різними методами (метод найменших квадратів, метод максимальної правдоподібності та ін.);
- перевірка статистичної значущості моделі.

Часто ***економетричну модель у загальному вигляді*** представляють як систему лінійних рівнянь:

$$BY=AX+E,$$

де B – матриця коефіцієнтів при ендогенних (залежних) змінних; Y – вектор ендогенних змінних (спостережуване значення залежної змінної); A – матриця коефіцієнтів при екзогенних (пояснюючих) змінних; X – вектор екзогенних змінних; AX – пояснена частина, що залежить від значень пояснюючих змінних; E – вектор випадкових збурень (помилок, відхилень).

Економетричні моделі включають досить широкий клас різних економіко-математичних моделей. Приведемо одну із **класифікацій економетричних моделей**.

1. ***За способом математичного подання*** економетричні моделі можна умовно розділити на прості й складні. *Прості* економетричні моделі представлені одним рівнянням, однією залежністю, *складні* – декількома рівняннями, декількома залежностями.

2. ***За кількістю факторних ознак,*** що включають у модель, прості економетричні моделі можна розділити на однофакторні й багатофакторні. *Однофакторні моделі* містять одну незалежну ознаку, *багатофакторні* – ряд незалежних ознак. Однофакторні й багатофакторні моделі можуть бути представлені *лінійними* та *нелінійними* функціями.

3. ***Складні економетричні моделі*** можуть бути представлені трьома видами систем одночасних рівнянь залежно від форми включення до правої частини ендогенних змінних. Звичайно виділяють три типи систем:

1) системи, що розв'язані відносно ендогенних змінних;

2) рекурсивні системи;

3) системи, що не розв'язані відносно ендогенних змінних.

4. ***Залежно від наявності (відсутності) у моделі фактора часу*** розрізняють *динамічні* й *статичні моделі.* Прикладами динамічних моделей виступають: *трендові моделі; моделі згладжування часових рядів; моделі декомпозиції часового ряду; авторегресійні моделі* й *моделі ковзної середньої; лагові моделі* та *динамічні регресійні моделі.*

Побудова економетричної моделі проводиться у кілька **основних етапів**:

1) якісний аналіз (постановка мети аналізу, визначення сукупності, визначення результативних і факторних ознак, вибір періоду, за який проводиться аналіз, вибір методу аналізу);

2) попередній аналіз сукупності, що моделюється (перевірка однорідності сукупності, виключення аномальних спостережень, уточнення необхідного обсягу ознак, установлення законів розподілу ознак);

3) побудова економетричної моделі (установлення переліку факторів, розрахунок оцінок параметрів рівнянь регресії, перебирання конкуруючих варіантів моделі);

4) оцінка адекватності моделі (перевірка статистичної істотності рівняння залежності в цілому та його окремих параметрів; перевірка відповідності формальних властивостей оцінок завданням дослідження);

5) економічна інтерпретація та практичне використання моделі.

Контрольні запитання

1. Сформулюйте визначення «математична модель».
2. У чому особливості економетричного моделювання?
3. Сформулюйте визначення «економетрична модель».
4. Назвіть етапи побудови економетричних моделей.
5. Які економетричні методи використовуються при побудові й аналізі економетричних моделей?

2. ПАРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

Якщо певному значенню величини X відповідає розподіл деякої величини Y, то між ними існує *статистична (стохастична, імовірнісна) залежність*.

Величина Y називається *залежною змінною, результуючим фактором (ознакою), ендогенною змінною*.

Величина X називається *незалежною змінною, екзогенною змінною*.

Залежність математичного очікування випадкової величини Y від значень змінної X називається *кореляційною залежністю*.

$$M_x(Y) = \varphi(x). \qquad (2.1)$$

Рівняння (2.1) називається рівнянням регресії, $\varphi(x)$ – функцією регресії, її графік – *лінією регресії*.

За результатами дослідів можна одержати:

$$\hat{y} = \hat{\varphi}(x, b_0, b_1, \ldots, b_p), \qquad (2.2)$$

де $\hat{y}$ – умовна (групова) середня змінної Y; $b_0, b_1, \ldots, b_p$ – параметри рівняння регресії.

Рівняння (2.2) називається *вибірковим рівнянням регресії*.

2.1. Лінійна парна регресія

Рівняння лінійної парної регресії має вигляд

$$\hat{y} = b_0 + b_1 x. \qquad (2.3)$$

Відповідно до **методу найменших квадратів (МНК)** невідомі параметри b_0 й b_1 вибираються таким чином, щоб сума квадратів відхилень емпіричних значень y_i від значень $\hat{y}_i$, що знайдені за рівнянням регресії (2.3), була мінімальною:

$$S = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2 \to \min. \qquad (2.4)$$

На підставі необхідної умови екстремуму функції двох змінних $S = S(b_0, b_1)$ (2.4) дорівнюємо до нуля її частинні похідні, тобто

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i) = 0; \\ \dfrac{\partial S}{\partial b_1} = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i)x_i = 0, \end{cases}$$

звідки після перетворень одержимо систему нормальних рівнянь для визначення параметрів лінійної системи:

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases} \tag{2.5}$$

Тепер, розділивши обидві частини рівняння (2.5) на n, одержимо систему нормальних рівнянь у вигляді:

$$\begin{cases} b_0 + b_1 \bar{x} = \bar{y}; \\ b_0 \bar{x} + b_1 \bar{x}^2 = \overline{xy}, \end{cases} \tag{2.6}$$

де відповідні середні визначаються за формулами

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n};$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}; \qquad \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}.$$

Підставляючи значення $b_0 = \bar{y} - b_1 \bar{x}$ з першого рівняння системи (2.6) у рівняння регресії (2.3), одержимо

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{K_{XY}}{s_x^2}, \tag{2.7}$$

де b_1 – вибірковий коефіцієнт регресії; K_{XY} – вибірковий кореляційний момент або вибіркова кореляція; s_x^2 – вибіркова дисперсія змінної X:

$$K_{XY} = \overline{xy} - \bar{x} \cdot \bar{y}, \; s_x^2 = \overline{x^2} - \bar{x}^2.$$

Приклад 2.1

Розглянемо залежність між змінним видобутком вугілля на одного робітника *Y* (т) і товщиною шару *X* (м) за даними десяти шахт (n = 10).

Таблиця 2.1

i	1	2	3	4	5	6	7	8	9	10
x_i	8	11	12	9	8	8	9	9	8	12
y_i	5	10	10	7	5	6	6	5	6	8

Будемо вважати, що залежність між змінними лінійна:

$\hat{y} = b_0 + b_1 x$ або $\hat{y} - \bar{y} = b_0 (x - \bar{x})$.

За табл. 2.1 обчислимо:

$$\sum_{i=1}^{n} x_i = 8 + 11 + 12 + 9 + 8 + 8 + 9 + 9 + 8 + 12 = 94,$$

$$\sum_{i=1}^{n} x_i^2 = 8^2 + 11^2 + 12^2 + 9^2 + 8^2 + 8^2 + 9^2 + 9^2 + 8^2 + 12^2 = 908,$$

$$\sum_{i=1}^{n} y_i = 5 + 10 + 10 + 7 + 5 + 6 + 6 + 5 + 6 + 8 = 68,$$

$$\sum_{i=1}^{n} x_i y_i = 8 \cdot 5 + 11 \cdot 10 + 12 \cdot 10 + 9 \cdot 7 + 8 \cdot 5 + 8 \cdot 6 + 9 \cdot 6 + 9 \cdot 5 + 8 \cdot 6 + 12 \cdot 8 = 664,$$

$$\bar{x} = \frac{94}{10} = 9{,}4; \quad \bar{y} = \frac{68}{10} = 6{,}8;$$

$$K_{XY} = \frac{664}{10} - 9{,}4 \cdot 6{,}8 = 2{,}48; \quad s_x^2 = \frac{908}{10} - 9{,}4^2 = 2{,}44; \quad b_1 = \frac{2{,}48}{2{,}44} = 1{,}016.$$

$$\hat{y} - 6{,}8 = 1{,}016 \cdot (x - 9{,}4); \quad \hat{y} = -2{,}75 + 1{,}016x.$$

Відзначимо, що вільний член рівняння ($b_0 = -2{,}75$) не має реального змісту.

Якщо значення змінних великі, то розрахунки можуть бути проведені у відхиленнях від середніх величин.

$$\begin{matrix} x_i' = x_i - \bar{x} \\ y_i' = y_i - \bar{y} \end{matrix}; \quad \hat{y}_i = b_1 \cdot x_i';$$

$$b_1 = \frac{\overline{x'y'}}{\overline{x'^2}};$$

$$\frac{\hat{y} - \bar{y}}{s_y} = \frac{b_1}{s_y} \cdot \frac{(x - \bar{x})}{s_x} \cdot s_x.$$

Величина $r = \dfrac{b_1 \cdot s_x}{s_y}$ називається *вибірковим коефіцієнтом кореляції*. Він показує, на скільки величин s_y зміниться *Y*, якщо *X* зміниться на одне s_x.

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x \cdot s_y}. \tag{2.8}$$

Підставивши у вираз (2.8) вихідні дані, одержимо:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \tag{2.9}$$

Коефіцієнт кореляції має наступні *властивості*:

1) він приймає значення на відрізку [–1; 1], тобто $-1 \le r \le 1$. Чим ближче |*r*| до 1, тим тісніше кореляційний зв'язок.

2) при $|r| = 1$ кореляційний зв'язок стає функціональним. При цьому всі спостережувані значення лежать на одній лінії.

3) при $r = 0$ кореляційний зв'язок відсутній та лінія регресії паралельна осі x.

При $r > 0$ ($b_1 > 0$) кореляційний зв'язок називається *прямим*.

При $r < 0$ ($b_1 < 0$) кореляційний зв'язок називається *зворотним*.

Приклад 2.2

За даними приклада 2.1 обчислити r.

$$\sum_{i=1}^{n} y_i^2 = 5^2 + 10^2 + 10^2 + 7^2 + 5^2 + 6^2 + 6^2 + 5^2 + 6^2 + 8^2 = 496.$$

$r = \dfrac{10 \cdot 664 - 94 \cdot 68}{\sqrt{10 \cdot 908 - 94^2} \cdot \sqrt{10 \cdot 496 - 68^2}} = 0{,}866,$ тобто зв'язок між змінними досить тісний.

Контрольні запитання

1. Які величини називаються екзогенними та ендогенними змінними?
2. Яка залежність називається кореляційною?
3. Що описує лінія регресії?
4. Наведіть алгоритм методу найменших квадратів.
5. У чому суть коефіцієнта кореляції, які він має властивості?

3. ОСНОВНІ ПОЛОЖЕННЯ РЕГРЕСІЙНОГО АНАЛІЗУ. ОЦІНКА ПАРАМЕТРІВ РЕГРЕСІЙНОЇ МОДЕЛІ. ТЕОРЕМА ГАУСА–МАРКОВА. ІНТЕРВАЛЬНА ОЦІНКА ФУНКЦІЇ РЕГРЕСІЇ ТА ЇЇ ПАРАМЕТРІВ

3.1. Основні положення регресійного аналізу

Як відзначено в розділі 2, розглянута в регресійному аналізі залежність Y від X може бути представлена у вигляді модельного рівняння регресії:

$$M_x(Y) = \varphi(X).$$

Окремі виміри величини Y будуть відрізнятися від обчислених значень за рахунок неврахованих факторів і помилок спостереження

$$Y = \varphi(X) + \varepsilon,$$

де ε – випадкова величина, що називається *збуренням або помилкою*.

Для лінійної моделі ці рівняння мають вигляд

$$M_x(Y) = \beta_0 + \beta_1 \cdot x. \qquad (3.1)$$

Спостережувані значення величини y визначаються за формулою:

$$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon. \qquad (3.2)$$

Основні передумови регресійного аналізу:

1. У моделі (3.2) збурення ε_i (або y_i) є випадковою величиною, а x_i – невипадковою величиною.

2. Математичне очікування збурення ε_i дорівнює нулю:

$$M(\varepsilon_i) = 0, \quad M(y_i) = \beta_0 + \beta_i x_i.$$

3. Дисперсія збурення ε_i (або y_i) є постійною величиною для будь-якого номера i:

$$D(\varepsilon_i) = \sigma^2 \text{ або } D(y_i) = \sigma^2.$$

Ця умова називається умовою *гомоскедастичності (рівномінливості)*.

4. Збурення ε_i та ε_j не корельовані:

$$M(\varepsilon_i \cdot \varepsilon_j) = 0, \text{ якщо } i \neq j.$$

5. Збурення ε_i (або y_i) мають нормальні розподіли.

3.2. Оцінка параметрів регресійної моделі

Вибірковою оцінкою рівняння (3.2) є рівняння регресії:

$$\hat{y} = b_0 + b_1 x.$$

Вплив неврахованих факторів і помилок спостережень у моделі (3.2) визначається за допомогою дисперсії збурень або залишкової дисперсії.

Незміщеною оцінкою залишкової дисперсії є:

$$s^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n - m} = \frac{\sum_{i=1}^{n} e_i^2}{n - 2}, \tag{3.3}$$

де y_i – спостережуване значення; $\hat{y}_i$ – групова середня, що знайдена за рівнянням регресії, $e_i = \hat{y}_i - y_i$ – вибіркова оцінка збурення ε_i або залишок регресії, $m = 2$ – число зв'язків (два рівняння для визначення b_0, b_1).

3.3. Теорема Гауса–Маркова

Теорема Гауса–Маркова. Якщо регресійна модель (3.2) задовольняє умовам 1–4 підрозділу 3.1, то оцінки b_0, b_1 мають найменшу дисперсію в класі всіх незміщених лінійних оцінок.

3.4. Інтервальна оцінка функції регресії та її параметрів

1. Знайдемо довірчий інтервал для умовного математичного очікування $M_x(Y)$ із заданою надійністю.

Припустимо, що γ – це задана надійність. Тоді мають місце наступні рівняння:

$\hat{y}-\bar{y}=b_1(x-\bar{x})$ – рівняння регресії.

$\hat{y}=\bar{y}+b_1(x-\bar{x})$.

$\sigma_{\hat{y}}^2=\sigma_{\bar{y}}^2+(x-\bar{x})^2\cdot\sigma_{b_1}^2$ – рівняння дисперсії.

Можна показати, що

$$\sigma_{\bar{y}}^2=\frac{\sigma^2}{n},\ \sigma_{b_1}^2=\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Тоді оцінкою дисперсії $\sigma_{\hat{y}}^2$ є:

$$s_{\hat{y}}^2=s^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right),$$

де s^2 – вибіркова оцінка σ^2.

Можна показати, що статистика $t=\dfrac{\hat{y}-M_x(Y)}{s_{\hat{y}}}$ розподілена за t-розподілом Ст'юдента з k ступенями вільності. За таблицею цього розподілу для рівня значущості α та числа ступенів вільності k знаходимо $t_{1-\alpha;\,k}$. Тоді довірчий інтервал для $M_x(Y)$ має вигляд:

$$\hat{y}_0-t_{1-\alpha;\,k}\cdot s_{\hat{y}_0}\le M_x(Y)\le\hat{y}_0+t_{1-\alpha;\,k}\cdot s_{\hat{y}_0},$$

де $s_{\hat{y}_0}^2=s^2\left(\dfrac{1}{n}+\dfrac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$.

Прогнозовані значення залежної змінної Y залежать від відхилення незалежного фактора осі $\bar{X}$. Чим більше це відхилення (останнє відхилення), тим більше помилка прогнозу.

2. Знайдемо довірчий інтервал для індивідуального значення y_0^*:

$$s^2_{y_0^*} = s^2_{\hat{y}} + s^2;$$

$$s^2_{y_0^*} = s^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right).$$

Довірчий інтервал для y_0^* на рівні значущості α та при кількості ступенів вільності *k* має вигляд:

$$\hat{y}_0 - t_{1-\alpha;\,k}\cdot s_{y_0^*} \le y_0^* \le \hat{y}_0 + t_{1-\alpha;\,k}\cdot s_{y_0^*}.$$

3. Знайти довірчий інтервал для параметра β_1.

Можна показати, що статистика $t = \frac{b_1 - \beta_1}{s}\cdot\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ має *t*-розподіл Ст'юдента. Тоді довірчий інтервал для параметра β_1 на рівні значущості α та при кількості ступенів вільності *k* має вигляд

$$b_1 - t_{1-\alpha;\,k}\cdot\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \le \beta_1 \le b_1 + t_{1-\alpha;\,k}\cdot\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

4. Знайти довірчий інтервал для дисперсії збурень σ^2.

Можна показати, що статистика $\frac{ns^2}{\sigma}$ має χ^2-розподіл з $k = n - 2$ ступенями вільності. Тоді довірчий інтервал для σ^2 на рівні значущості α має вигляд

$$\frac{ns^2}{\chi^2_{\alpha/2;\,n-2}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1\text{-}\alpha/2;\,n\text{-}2}}.$$

Приклад 3.1

За даними приклада 2.1 знайти наступні величини:

1. Довірчий інтервал для $M_x(Y)$ при $X_0 = 8$.

а) $\hat{y}_{X_0} = \hat{y}_0 = -2{,}75 + 1{,}016 \cdot 8 = 5{,}38.$

б) $\sum\limits_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 24{,}4.$

$$\hat{y}_i = -2{,}75 + 1{,}016 \cdot x_i \text{ (знайти 10 значень } \hat{y}_i\text{)}.$$

$$\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 = 8{,}39.$$

$$s^2 = \frac{8{,}39}{10-2} = 1{,}049.$$

$$s^2_{\hat{y}_{x_0}} = 1{,}049 \cdot \left(\frac{1}{10} + \frac{(8-9{,}4)^2}{24{,}4}\right) = 0{,}189.$$

$$s_{\hat{y}_{x_0}} = \sqrt{0{,}189} = 0{,}435.$$

в) за таблицею t-розподілу при α = 0,05 та k = 8 значення статистики дорівнює $t_{0{,}95;\,8}$ = 2,31. Тоді знаходимо довірчий інтервал для $M_x(Y)$:

$$5{,}38 - 2{,}31 \cdot 0{,}435 \le M_x(Y) \le 5{,}38 + 2{,}31 \cdot 0{,}435;$$

$$4{,}38 \le M_x(Y) \le 6{,}38.$$

2. Знайдемо довірчий інтервал для y_0^* при α = 0,05 та k = 8:

$$s^2_{y_0^*} = 0{,}189 + 1{,}049 = 1{,}238.$$

$$s_{y_0^*} = 1{,}113.$$

$$5{,}38 - 2{,}31 \cdot 1{,}113 \le y_0^* \le 5{,}38 + 2{,}31 \cdot 1{,}113;$$

$$2{,}81 \le y_0^* \le 7{,}95.$$

3. Знайдемо довірчий інтервал для β_1 при α = 0,05 та k = 8:

$$1{,}016 - 2{,}31 \cdot \frac{\sqrt{1{,}049}}{\sqrt{24{,}4}} \le \beta_1 \le 1{,}016 + 2{,}31 \cdot \frac{\sqrt{1{,}049}}{\sqrt{24{,}4}};$$

$$0{,}537 \le \beta_1 \le 1{,}495.$$

Це означає, що при зміні товщини шару на 1 м добовий видобуток вугілля буде змінюватися від 0,537 до 1,495 (т).

4. Знайдемо довірчий інтервал для σ^2 при $\alpha = 0,05$ та $k = 8$:

$$\chi^2_{\alpha/2;\,k} = \chi^2_{0,025;\,8} = 17,53;$$

$$\chi_{1-\alpha/2;\,k} = \chi_{0,975;\,8} = 2,18.$$

$$\frac{10 \cdot 1,049}{17,53} \le \sigma^2 \le \frac{10 \cdot 1,049}{2,18};$$

$$0,598 \le \sigma^2 \le 4,81.$$

$$0,773 \le \sigma \le 2,19.$$

Таким чином, з надійністю 0,95 дисперсія збурення знаходиться в межах від 0,598 до 4,81, а їхнє стандартне відхилення – від 0,773 до 2,19 (тонн).

Контрольні запитання

1. Перелічіть основні передумови регресійного аналізу.
2. У чому суть теореми Гауса–Маркова?
3. Що характеризує *t*-критерій Ст'юдента? Як визначається статистична значущість оцінок параметрів моделі?
4. Як визначаються довірчі інтервали для оцінок параметрів моделі?

4. ОЦІНКА ЗНАЧУЩОСТІ РІВНЯННЯ РЕГРЕСІЇ. КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ. РАНГОВА КОРЕЛЯЦІЯ. КОЕФІЦІЄНТ СПІРМЕНА

4.1. Оцінка значущості рівняння регресії. Коефіцієнт детермінації

Перевірити значущість рівняння регресії – значить установити, чи відповідає математична модель експериментальним даним і чи досить включених у рівняння пояснюючих змінних (однієї або декількох) для опису результуючих факторів.

Перевірка значущості рівняння регресії може бути проведена способом оцінки значущості коефіцієнта регресії b_1, що має t-розподіл з k ступенями вільності.

Рівняння парної регресії (або коефіцієнта регресії b_1) є значущім на рівні α, якщо спостережуване значення статистики $t = \frac{b_1}{s}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ за абсолютною величиною більше табличного значення $t_{1-\alpha;\, k}$.

Приклад 4.1

Перевірити значущість рівняння регресії в прикладі 2.1 на рівні значущості $\alpha = 0{,}05$; $k = 8$.

$$t_{0,95;\,8} = 2{,}31.$$

$$t = \frac{1{,}016}{\sqrt{1{,}049}}\sqrt{24{,}4} = 4{,}9.$$

Оскільки $t = 4{,}9 > 2{,}31 = t_{0,95;\,8}$, то коефіцієнт регресії b_1, а також рівняння регресії значущі.

Обчислимо *суму квадратів відхилень*.

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}[(\hat{y}_i-\bar{y})+(y_i-\hat{y}_i)]^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+(y_i-\hat{y}_i)^2+2(\hat{y}_i-\bar{y})(y_i-\hat{y}_i).$$

Можна показати, що третій доданок дорівнює нулю, тоді це рівняння запишемо так:

$$Q=Q_R+Q_e,$$

де Q – загальна сума квадратів відхилень залежної змінної від середньої; Q_R – сума квадратів, що обумовлена регресією; Q_e – сума квадратів, що обумовлена неврахованими факторами.

Ефективною оцінкою адекватності регресійної моделі є *коефіцієнт детермінації,* який обчислюється за допомогою наступної формули:

$$R^2=\frac{Q_R}{Q}=1-\frac{Q_e}{Q}.$$

Коефіцієнт детермінації показує, яка частина зміни залежної змінної обумовлена зміною незалежного фактора.

$$0\le R^2\le 1.$$

Чим ближче R^2 до 1, тим краще рівняння регресії наближає (апроксимує) експериментальні дані.

У випадку парної регресії $R^2=r^2$, де r – коефіцієнт кореляції.

Приклад 4.2

За даними приклада 2.1 знайти коефіцієнт детермінації.

$$Q_R=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=25{,}21.$$

$$Q=\sum_{i=1}^{n}(y_i-\bar{y})^2=33{,}6.$$

$$R^2=\frac{25{,}21}{33{,}6}=0{,}75.$$

У прикладі 3.1 було знайдено $r = 0{,}886; \quad r^2 = (0{,}886)^2 = 0{,}75.$

Висновок: зміна залежної змінної (змінного видобутку вугілля) на 75 % обумовлена зміною товщини шару вугілля.

4.2. Рангова кореляція. Коефіцієнт Спірмена

Якщо між двома змінними, що виражені своїми рангами, існує кореляційний зв'язок, то він називається *ранговою кореляцією*. Тісноту такого зв'язку визначає **коефіцієнт рангової кореляції Спірмена**:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n},$$

де r_i, s_i – ранги i-го об'єкту за двома показниками; n – число спостережень.

Приклад 4.3

Десять студентів склали іспити за двома дисциплінами A і B. На основі набраних балів отримані наступні ранги (табл. 4.1). Обчислити коефіцієнт рангової кореляції Спірмена й перевірити його значущість на рівні $\alpha = 0{,}05$.

Таблиця 4.1

A спадне	95	90	86	84	75	70	62	60	57	50
B випадкове	92	93	83	80	55	60	45	72	62	70
$r_{i(A)}$	1	2	3	4	5	6	7	8	9	10
$s_{i(B)}$	2	1	3	4	9	8	10	5	7	6

$$\sum_{i=1}^{n}(r_i - s_i)^2 = (1-2)^2 + (2-1)^2 + (3-3)^2 + (4-4)^2 + (5-9)^2 + (6-8)^2 + (7-10)^2 +$$
$$+(8-5)^2 + (9-7)^2 + (10-6)^2 = 60.$$

$$\rho_B = 1 - \frac{6 \cdot 60}{10^3 - 10} = 0{,}64.$$

Для визначення значущості цього коефіцієнта (ρ_B) обчислимо статистику:

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}};$$

$$t = \frac{0{,}64\sqrt{10-2}}{\sqrt{1-0{,}64^2}} = 2{,}36.$$

За таблицею t-розподілу на рівні значущості $\alpha = 0{,}05$ та числа ступенів вільності $n = 10–2 = 8$ одержуємо: $t_{кр} = 2{,}31$.

Оскільки $t = 2{,}36 > 2{,}31 = t_{кр}$, то оцінка ρ_B значуща, тобто між двома результатами іспитів існує тісний зв'язок.

Контрольні запитання

1. Навіщо необхідно перевіряти статистичну значущість оцінок параметрів моделі?
2. Що показує коефіцієнт детермінації, у яких межах він змінюється?
3. Що визначає коефіцієнт рангової кореляції Спірмена?

5. МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

5.1. Елементи лінійної алгебри

Матрицею розміру $(m \times n)$ називається таблиця чисел, що містить m рядків та n стовпців:

$$A_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Дії з матрицями:

1. *Множення матриць.*

$$C_{m\times l} = A_{m\times n} \cdot B_{p\times l}.$$

1.1. Множити можна тільки узгоджені матриці (треба, щоб $n = p$).

1.2. Розмір матриці-добутку буде дорівнювати $m \times l$.

1.3. $C_{ij} = \sum_{k=1}^{n} a_{ik} \cdot b_{kj}$.

Приклад 5.1

$$A_{2\times 3} = \begin{pmatrix} 1 & 2 & 3 \\ -3 & 1 & 2 \end{pmatrix}; \; B_{3\times 2} = \begin{pmatrix} 2 & -1 \\ 3 & 4 \\ -1 & 2 \end{pmatrix}.$$

$$C_{2\times 2} = A_{2\times 3} \cdot B_{3\times 2} = \begin{pmatrix} 1 & 2 & 3 \\ -3 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 3 & 4 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 13 \\ -5 & 11 \end{pmatrix}.$$

$$C_{11} = 1\cdot 2 + 2\cdot 3 + 3\cdot(-1) = 5;$$

$$C_{12} = 1\cdot(-1) + 2\cdot 4 + 3\cdot 2 = 13;$$

$$C_{21} = (-3)\cdot 2 + 1\cdot 3 + 2\cdot(-1) = -5;$$

$$C_{22} = -3\cdot(-1) + 1\cdot 4 + 2\cdot 2 = 11.$$

2. *Знаходження зворотної матриці.*

2.1. Зворотна матриця існує тільки для невиродженої квадратної матриці (якщо $\Delta A = 0$, то матриця вироджена, якщо $\Delta A \neq 0$ – невироджена); $m = n$ – квадратна матриця.

2.2. *Алгоритм знаходження зворотної матриці*

$$A^{-1} \cdot A = E,$$

де E – одинична матриця до матриці A.

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

1) Знайти *визначник матриці* ΔA.

2) Знайти *транспоновану матрицю* A'.

3) Знайти *приєднану матрицю* A^S (її елементами є алгебричні доповнення відповідних елементів матриці A').

4) $A^{-1} = \dfrac{1}{\Delta A} \cdot A^S$.

Приклад 5.2

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 3 \\ 1 & 4 & -2 \end{pmatrix}.$$

1) $\Delta A = 6 \neq 0$.

2) $A' = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 4 \\ -1 & 3 & -2 \end{pmatrix}$.

3) $A^S = \begin{pmatrix} -12 & 0 & 6 \\ 7 & -1 & -5 \\ 8 & -2 & -4 \end{pmatrix}$.

4) $A^{-1}=\frac{1}{6}\cdot\begin{pmatrix} -12 & 0 & 6 \\ 7 & -1 & -5 \\ 8 & -2 & -4 \end{pmatrix}.$

5) $\frac{1}{6}\cdot\begin{pmatrix} -12 & 0 & 6 \\ 7 & -1 & -5 \\ 8 & -2 & -4 \end{pmatrix}\begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 3 \\ 1 & 4 & -2 \end{pmatrix}=E.$

5.2. Класична нормальна лінійна модель

Нехай *Y* – залежна змінна. Тоді змінні фактори, від яких залежить величина *Y*, будемо позначати: $X_1, X_2, \ldots, X_p$.

Припустимо, що зв'язок між цими змінними лінійний, тоді *модель множинної регресії* має вигляд

$$y_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots+\beta_p x_{ip}+\varepsilon_i, \tag{5.1}$$

де $i=\overline{1,n}$; y_i – залежна змінна; x_i – i-те спостереження відповідних факторів.

Ця модель задовольняє основним передумовам регресійного аналізу (розділ 3).

Введемо наступні позначення:

1) $Y=\left(y_1, y_2, \ldots y_n\right)'$ називається *матрицею-стовпцем* (або вектором) залежної змінної *Y*.

2) $X=\begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix}$ називається *матрицею значень факторів* або *матрицею плану* розміру $n\times(p+1)$.

3) $\beta = (\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_p)'$ – *матриця-стовпець (вектор) параметрів.*

4) $\varepsilon = (\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_p)'$ – *матриця-стовпець (вектор) збурень.*

Тоді система рівнянь (5.1) у матричній формі має вигляд

$$Y = X\beta + \varepsilon. \tag{5.2}$$

Вибірковою оцінкою цієї моделі є

$$Y = Xb + e, \tag{5.3}$$

де $b = (b_0 \quad b_1 \quad \dots \quad b_p)'$, $e = (e_1 \quad e_2 \quad \dots \quad e_n)'$.

Для знаходження вектора оцінок параметрів b застосуємо МНК. Сума квадратів відхилень дорівнює:

$$S = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} e_i^2 .$$

$$\sum_{i=1}^{n} e_i^2 = (e_1 \quad e_2 \quad \dots \quad e_n) \cdot (e_1 \quad e_2 \quad \dots \quad e_n)'.$$

$$\sum_{i=1}^{n} e_i^2 = (Y - Xb) \cdot (Y - Xb)' \to \min .$$

Остаточно маємо: $S = (Y - Xb) \cdot (Y - Xb)'$.

Можна показати, що вектор частинних похідних дорівнює

$$\frac{\partial S}{\partial b} = -2X'Y + 2X'Xb.$$

Дорівнявши похідну до нуля, одержуємо систему нормальних рівнянь у матричній формі для знаходження вектора b.

$$X'Xb = X'Y. \tag{5.4}$$

Розв’язком цього матричного рівняння є вектор:

$$b = (X'X)^{-1} X'Y. \tag{5.5}$$

Рівняння (5.4) справедливо при виконанні наступних *умов*:

1) вектор збурень ε – випадковий вектор, X – невипадкова матриця;

2) $M(\varepsilon) = 0_n$;

3, 4) $\Sigma_\varepsilon = M(\varepsilon\varepsilon') = \sigma^2 E_n$;

5) ε – нормально розподілений випадковий вектор;

6) щоб існувала зворотна матриця $(X'X)^{-1}$, необхідна умова:

$$\Delta(X'X)^{-1} \neq 0.$$

У цьому випадку ранг $r(X) = p + 1 < n$, де n – число спостережень, що повинне перевищувати ранг для надійності одержуваних результатів.

Модель, що задовольняє цим умовам, називається ***класичною нормальною лінійною моделлю множинної регресії***. Якщо не виконується умова 5, то в цій назві викидається слово «нормальна».

При відомому векторі b рівняння множинної регресії має вигляд

$$\hat{y}_0 = X_0' b,$$

де $X_0' = (1 \quad x_{10} \quad x_{20} \quad \dots \quad x_{p0})$.

Приклад 5.3

Є дані про змінний видобуток вугілля на одного робітника Y (т), товщину шару X_1 (м) та рівень механізації робіт X_2 (%) (табл. 5.1). Припускаючи, що між цими змінними існує лінійна залежність, знайти рівняння множинної регресії.

Таблиця 5.1

i	x_{i1}	x_{i2}	y_i	x_{i1}^2	x_{i2}^2	y_i^2	$x_{i1}x_{i2}$	$x_{i1}y_i$	$x_{i2}y_i$	$\hat{y}_i$	$e_i^2 = \left(\hat{y}_i - y_i\right)^2$
1	8	5	5	64	25	25	40	40	25	5,13	0,016
2	11	8	10	121	64	100	88	110	80	8,79	1,464
3	12	8	10	144	64	100	96	120	80	9,64	1,127
4	9	5	7	81	25	49	45	63	35	5,98	1,038
5	8	7	5	64	49	25	56	40	35	5,86	0,741
6	8	8	6	64	64	36	64	48	48	6,23	0,052
7	9	6	6	81	36	36	54	54	36	6,35	0,121
8	9	4	5	81	16	25	36	45	20	5,61	0,377
9	8	5	6	64	25	36	40	48	30	5,13	0,762
10	12	7	8	144	49	64	84	96	56	9,28	1,631
Σ	94	63	68	908	417	496	603	664	445	–	6,329

Позначимо

$$Y = \begin{pmatrix} 5 \\ 10 \\ \dots \\ 8 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{pmatrix}.$$

$$X'X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 8 & 11 & \dots & 12 \\ 5 & 8 & \dots & 7 \end{pmatrix} \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{pmatrix} = \begin{pmatrix} 10 & 94 & 63 \\ 94 & 908 & 603 \\ 63 & 603 & 417 \end{pmatrix}.$$

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 8 & 11 & \dots & 12 \\ 5 & 8 & \dots & 7 \end{pmatrix} \begin{pmatrix} 5 \\ 10 \\ \dots \\ 8 \end{pmatrix} = \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix}.$$

Знайдемо $\left(X'X\right)^{-1}$:

1) $\Delta\left(X'X\right) = 3738 \neq 0.$

2) транспонуємо $X'X$:

$$\left(X'X\right)' = X'X.$$

3) знайдемо приєднану матрицю $\left(X'X\right)^{S}$:

$$\left(X'X\right)^{S} = \begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}.$$

4) знайдемо зворотну матрицю:

$$\left(X'X\right)^{-1} = \frac{1}{\Delta}\left(X'X\right)^{S}$$

$$\left(X'X\right)^{-1} = \frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}.$$

За формулою (5.5) обчислюємо значення коефіцієнта b:

$$b = \frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}\begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix} = \begin{pmatrix} -3{,}54 \\ 0{,}854 \\ 0{,}364 \end{pmatrix}.$$

Рівняння регресії має вигляд:

$$\hat{y} = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix}\begin{pmatrix} -3{,}54 \\ 0{,}854 \\ 0{,}364 \end{pmatrix}.$$

$$\hat{y} = -3{,}54 + 0{,}854x_1 + 0{,}367x_2.$$

Це рівняння показує, що при збільшенні товщини шару X_1 (X_2 = const) на 1 м видобуток вугілля збільшується на 0,854 т, а при збільшенні тільки рівня механізації X_2 (X_1 = const) на 1 % видобуток вугілля збільшується на 0,367 т.

Для порівняння впливу на залежну змінну різних факторів використовують стандартизовані коефіцієнти регресії b_j' та еластичності E_j' $\left(j=\overline{1,\ p}\right)$.

$$b_j' = b_j \frac{S_{x_j}}{S_y};$$

$$E_j' = b_j \frac{\overline{x}_j}{\overline{y}}.$$

$$S_{X_j} = \overline{x_j^2} - \left(\overline{x}\right)^2; \quad S_y = \overline{y^2} - \left(\overline{y}\right)^2.$$

Приклад 5.4

За даними приклада 5.3 порівняти роздільний вплив на видобуток вугілля двох факторів: товщини шару й рівня механізації.

Раніше було обчислено

$$\overline{x}_1 = 9{,}4; \quad \overline{y} = 6{,}8; \quad S_{X_1} = 1{,}56.$$

Обчислимо додатково:

$$\overline{x}_2 = 6{,}3; \quad S_{X_2} = 1{,}42; \quad S_y = 1{,}83; \quad b_1 = 0{,}854; \quad b_2 = 0{,}367.$$

$$b_1' = 0{,}854 \frac{1{,}56}{1{,}83} = 0{,}728;$$

$$b_2' = 0{,}367 \frac{1{,}42}{1{,}83} = 0{,}285;$$

$$E_1' = 0{,}854 \frac{9{,}4}{6{,}8} = 1{,}18;$$

$$E_2' = 0{,}367 \frac{6{,}3}{6{,}8} = 0{,}34.$$

Висновок: за обома показниками на видобуток вугілля більше впливає фактор товщини шару, ніж фактор рівня механізації.

Перетворимо вектор оцінок (5.5). Маємо

$$b = (X'X)^{-1} X'Y .$$

$$Y = X\beta + \varepsilon .$$

$$b = (X'X)^{-1} X'(X\beta + \varepsilon) = (X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon =$$

$$= E\beta + (X'X)^{-1} X'\varepsilon = \beta + (X'X)^{-1} X'\varepsilon .$$

Оцінка параметра *b* містить випадкові помилки.

$M(b) = M(\beta) + M\left[(X'X)^{-1} X'\varepsilon\right].$

$M(b) = \beta + (X'X)^{-1} X'M(\varepsilon).$

$M(\varepsilon) = 0$ – за умовами математичної моделі.

$M(b) = \beta.$

5.3. Коваріаційна матриця

Коваріаційна матриця визначає розсіювання оцінок рівняння регресії. У загальному вигляді вона записується так:

$$\Sigma_b = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \dots & \sigma_{0p} \\ \sigma_{10} & \sigma_{11} & \dots & \sigma_{1p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p0} & \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix}, \tag{5.6}$$

де $\sigma_{ij} = M\left[(b_i - M(b_i))(b_j - M(b_j))\right]$ – *коваріації* (або *кореляційні моменти* двох змінних b_i, b_j).

Коваріація показує ступінь розсіювання цих змінних та їхній взаємозв'язок.

$$\sigma_{ij} = M\left[(b_i - \beta_i)(b_j - \beta_j)\right].$$

На головній діагоналі елементи матриці (5.6):

$$\sigma_{jj} = \sigma_{b_j} = M\left[(b_j - \beta_j)^2\right].$$

Тоді матрицю (5.6) можна представити:

$$\Sigma_b = M\left[(b-\beta)(b-\beta)'\right].$$

$$b = \beta + \left(X'X\right)^{-1} X'\varepsilon .$$

$$\Sigma_b = M\left[\left(\left(X'X\right)^{-1} X'\varepsilon\right)\left(\left(X'X\right)^{-1} X'\varepsilon\right)'\right] = M\left[\left(X'X\right)^{-1} X'\varepsilon\varepsilon' X\left(X'X\right)^{-1}\right] =$$

$$= \left(X'X\right)^{-1} X' M(\varepsilon\varepsilon') X\left(X'X\right)^{-1}.$$

Матриця $M(\varepsilon\varepsilon')$ являє собою коваріаційну матрицю вектора збурень ε.

$$M(\varepsilon\varepsilon') = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_n) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_n) \\ \dots & \dots & \dots & \dots \\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \dots & M(\varepsilon_n^2) \end{pmatrix}. \qquad (5.7)$$

За умовами 3, 4 підрозділу 5.2:

$$M(\varepsilon\varepsilon') = \sigma^2 E_n,$$

де E_n – одинична матриця n-го порядку.

$$\Sigma_b = \left(X'X\right)^{-1} X'\sigma^2 E_n X\left(X'X\right)^{-1} = \sigma^2\left(X'X\right)^{-1}.$$

Матриця $\left(X'X\right)^{-1}$ визначає не тільки вектор b, але й дисперсії його компонентів.

Контрольні запитання

1. Яка необхідна умова існування зворотної матриці?

2. Наведіть алгоритм методу найменших квадратів для оцінки параметрів множинної лінійної моделі.

3. Яким умовам повинна задовольняти класична нормальна лінійна модель множинної регресії?

4. Що визначає коваріаційна матриця, у чому суть коваріації?

6. ДОВІРЧІ ІНТЕРВАЛИ ДЛЯ КОЕФІЦІЄНТІВ І ФУНКЦІЇ РЕГРЕСІЇ. ОЦІНКА ЗНАЧУЩОСТІ РІВНЯННЯ МНОЖИННОЇ РЕГРЕСІЇ

6.1. Довірчі інтервали для коефіцієнтів і функції регресії

Оцінкою $s_{b_j}^2$ дисперсії $\sigma_{b_j}^2$ є величина:

$$s_{b_j}^2 = s^2 \left[\left(X'X \right)^{-1} \right]_{jj},$$

де $s^2 = \dfrac{\sum_{i=1}^{n} e_i^2}{n - p - 1}$ – незміщена оцінка σ^2; $\left[\left(X'X \right)^{-1} \right]_{jj}$ – діагональний елемент зворотної матриці.

Коефіцієнт b_j є значущим на рівні α, якщо

$$|t| = \frac{|b_j|}{s_{b_j}} > t_{1-\alpha;\,k}.$$

Тоді довірчий інтервал параметра β_j має вигляд

$$b_j - t_{1-\alpha;\,k} \cdot s_{b_j} \le \beta_j \le b_j + t_{1-\alpha;\,k} \cdot s_{b_j}.$$

Знайдемо довірчий інтервал для математичного очікування $M_{X_0}(Y)$ за умови, що вектор *X* прийняв значення:

$$X_0' = \begin{pmatrix} 1 & x_{10} & x_{20} & \dots & x_{p0} \end{pmatrix}.$$

Знаходимо з рівняння регресії групову середню величину, що ділена на $\hat{y}_0$.

$$s_{\hat{y}_0} = s\sqrt{X_0' \left(X'X \right)^{-1} X_0}.$$

Тоді довірчий інтервал має вигляд

$$\hat{y}_0 - t_{1-\alpha;\,k} \cdot s_{\hat{y}_0} \le M_X(Y) \le \hat{y}_0 + t_{1-\alpha;\,k} \cdot s_{\hat{y}_0}.$$

Знайдемо довірчий інтервал для індивідуального значення y_0^*:

$$s^2_{y_0^*} = s^2_{\hat{y}_0} + s^2.$$

$$\hat{y}_0 - t_{1-\alpha;\,k} \cdot s_{y_0^*} \le y_0^* \le \hat{y}_0 + t_{1-\alpha;\,k} \cdot s_{y_0^*}.$$

Знайдемо довірчий інтервал для дисперсії σ^2:

$$\frac{ns^2}{\chi_{\frac{\alpha}{2};\,k}} \le \sigma^2 \le \frac{ns^2}{\chi_{1-\frac{\alpha}{2};\,k}}.$$

Приклад 6.1

За даними прикладу 5.3 оцінити видобуток вугілля для шару товщини $x_1 = 8$ м і рівня механізації робіт 6 %.

$$X_0' = \begin{pmatrix} 1 & 8 & 6 \end{pmatrix}'.$$

$$\hat{y}_0 = -3{,}54 + 0{,}854 x_1 + 0{,}367 x_2;$$

$$\hat{y}_0 = -3{,}54 + 0{,}854 \cdot 8 + 0{,}367 \cdot 6 = 5{,}49.$$

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} = \frac{6{,}32}{10-2-1} = 0{,}904.$$

$$s = \sqrt{0{,}904} = 0{,}951.$$

$$X_0'(X'X)^{-1}X_0 = \begin{pmatrix} 1 & 8 & 6 \end{pmatrix} \frac{1}{3738} \begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix} \begin{pmatrix} 1 \\ 8 \\ 6 \end{pmatrix} = 0{,}187.$$

$$s_{\hat{y}_0} = 0{,}951\sqrt{0{,}187} = 0{,}411.$$

За таблицею знаходимо $t_{0,95;\,7} = 2{,}36.$

Визначимо довірчий інтервал для $M_{X_0}(Y)$:

$$5{,}49 - 2{,}36 \cdot 0{,}411 \le M_{X_0}(Y) \le 5{,}49 + 2{,}36 \cdot 0{,}411;$$

$$4{,}52 \le M_{X_0}(Y) \le 6{,}46.$$

Висновок: порівнюючи отриманий результат із прикладом 3.1, відзначимо зменшення величини довірчого інтервалу $M_{X_0}(Y)$. Це пов'язано з тим, що включення нового фактора x_2 підвищило точність рівняння множинної регресії.

Знайдемо довірчий інтервал для індивідуального значення y_0^*:

$$s_{y_0^*}^2 = s_{\hat{y}_0}^2 + s^2 = (0{,}951)^2 + (0{,}951)^2 \cdot 0{,}187 = (1{,}036)^2;$$

$$s_{y_0^*} = 1{,}036.$$

$$5{,}49 - 2{,}36 \cdot 1{,}036 \le y_0^* \le 5{,}49 + 2{,}36 \cdot 1{,}036;$$

$$3{,}05 \le y_0^* \le 7{,}93.$$

Знайдемо довірчий інтервал і перевіримо значущість коефіцієнтів регресії b_1, b_2.

$$s_{b_1} = 0{,}951 \cdot \sqrt{\frac{1}{3738} 201} = 0{,}221;$$

$$s_{b_2} = 0{,}951 \cdot \sqrt{\frac{1}{3738} 244} = 0{,}243.$$

Обчислимо *t*-вибіркове:

$$t_{b_1} = \frac{0{,}854}{0{,}221} = 3{,}81; \quad t_{0{,}95;\,7} = 2{,}36.$$

Оскільки $t_{b_1} > 2{,}36$, то коефіцієнт b_1 є значущим на цьому рівні.

$$t_{b_2} = \frac{0{,}367}{0{,}243} = 1{,}51.$$

Далі, оскільки $t_{b_2} < 2{,}36$, то b_2 – не є значущим на рівні $\alpha = 0{,}05$.

Значить, знаходити довірчий інтервал для b_2 не має реального змісту.

Довірчий інтервал для β_1:

$$0{,}854 - 2{,}36 \cdot 0{,}21 \le \beta_1 \le 0{,}854 + 2{,}36 \cdot 0{,}21;$$

$$0{,}332 \le \beta_1 \le 1{,}376.$$

Знайдемо довірчий інтервал для дисперсії σ^2 на рівні $\alpha = 0{,}05$:

$$\chi^2_{0{,}025;\,7} = 16{,}01; \quad \chi^2_{0{,}975;\,7} = 1{,}69.$$

$$\frac{10 \cdot 0{,}904}{16{,}01} \le \sigma^2 \le \frac{10 \cdot 0{,}904}{1{,}69};$$

$$0{,}565 \le \sigma^2 \le 5{,}349;$$

$$0{,}751 \le \sigma \le 2{,}313.$$

6.2. Оцінка значущості рівняння множинної регресії

Обчислимо *коефіцієнт детермінації множинної регресії R^2*.

Сума загального числа квадратів відхилень дорівнює:

$$Q = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}.$$

$$\sum_{i=1}^{n} y_i^2 = Y'Y.$$

$$Q = Y'Y - n\bar{y}^2.$$

Суми квадратів відхилень, що обумовлені неврахованими факторами Q_e й регресією Q_R, дорівнюють:

$$Q_e = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = e'e = \left(Y - Xb\right)'\left(Y - Xb\right) = \left|X'Xb = X'Y\right| = Y'Y - b'X'Y.$$

$$Q_R = Q - Q_e = b'X'Y - n\left(\bar{y}\right)^2.$$

Тоді *коефіцієнт детермінації* розраховується за формулою

$$R^2 = Q_R / Q = \frac{b'X'Y - n\left(\bar{y}\right)^2}{Y'Y - n\left(\bar{y}\right)^2}. \qquad (6.1)$$

Недоліком коефіцієнта детермінації (6.1) є те, що він збільшується при введенні нових факторів, хоча це не завжди означає поліпшення регресійної моделі.

У цьому випадку краще використати *скоректований коефіцієнт детермінації*:

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}\left(1 - R^2\right). \qquad (6.2)$$

За відомим коефіцієнтом детермінації визначимо *значущість рівняння регресії*. Обчислимо *статистику*:

$$F = \frac{R^2(n-p-1)}{\left(1-R^2\right)p}, \qquad (6.3)$$

де p – число параметрів.

За таблицею *Фішера–Снедекора* знаходимо $F_{\alpha;\,p;\,n-p-1}$, якщо $F > F_{\alpha;\,p;\,n-p-1}$, то рівняння регресії значуще.

Приклад 6.2

За даними прикладу 5.3 визначити R^2 і перевірити значущість рівняння регресії на рівні 0,05.

$$Y'Y = \sum_{i=1}^{n} y_i^2 = 496.$$

$$b'X'Y = \left(-3{,}54 \quad 0{,}854 \quad 0{,}367\right)\begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix} = 489{,}65.$$

$$\bar{y} = 6{,}8.$$

За формулою (6.1) $\quad R^2 = \dfrac{489{,}65 - 10 \cdot (6{,}8)^2}{496 - 10 \cdot (6{,}8)^2} = 0{,}811.$

$$F = \frac{0{,}811(10-2-1)}{(1-0{,}811)2} = 15.$$

За таблицею *Фішера–Снедекора* знаходимо $F_{0,05;\,2;\,7} = 4{,}74$.

Оскільки $F > F_{0,05;\,2;\,7}$; $15 > 4{,}74$, то рівняння регресії значуще. Значущість означає, що включені в модель параметри x_1 і x_2 досить добре визначають регресійну модель.

Контрольні запитання

1. Як визначаються довірчі інтервали для коефіцієнтів і функції множинної регресії?
2. Що показує коефіцієнт детермінації множинної регресії, у яких межах він змінюється?
3. Що характеризує F-критерій Фішера в кореляційному аналізі?
4. Що показує рівень значущості в F-критерії Фішера?

7. ДЕЯКІ ОСОБЛИВОСТІ ВИКОРИСТАННЯ РЕГРЕСІЙНОЇ МОДЕЛІ

7.1. Мультиколінеарність

Мультиколінеарністю називається висока залежність (корельованість) різних факторів.

Якщо між факторами є тісний зв'язок, то визначник матриці $\left(X'X\right)$ дуже малий. Це призводить до великих помилок при оцінці параметрів рівняння регресії. В остаточному підсумку виходить незначуще рівняння регресії, що не має реального змісту.

Усунути або зменшити *мультиколінеарність* можна **методом покрокового відбору найбільш інформативних факторів**. Цю процедуру розглянемо на прикладі.

Приклад 7.1

За даними $n = 20$ сільськогосподарських районів урожайність Y (ц/га) залежіть від факторів:

X_1 – число тракторів на 100 га;

X_2 – число комбайнів на 100 га;

X_3 – число знарядь обробки ґрунту на 100 га;

X_4 – число добрив на 1 га (т/га);

X_5 – кількість засобів хімічного захисту на 1 га (ц/га).

Таблиця 7.1

i	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
1	9,7	1,59	0,26	2,05	0,32	0,14
2	8,4	0,34	0,28	0,46	0,59	0,66
…	…	…	…	…	…	…
20	8,7	1,36	0,26	0,17	0,99	0,42

За даними табл. 7.1 складемо рівняння регресії:

$$\hat{y} = 3{,}515 - 0{,}006x_1 + 15{,}5x_2 + 0{,}11x_3 + 4{,}475x_4 - 2{,}932x_5.$$
$$(5{,}41) \quad (0{,}6) \qquad (21{,}59) \quad (0{,}85) \quad (1{,}54) \qquad (3{,}09)$$

Примітка. У дужках наведені середньоквадратичні відхилення.

Визначимо значущість коефіцієнтів рівняння регресії:

$$t_{b_j} = \frac{|b_j|}{S_{b_j}}.\ t_{b_0} = 0{,}65;\ t_{b_1} = 0{,}01;\ t_{b_2} = 0{,}72;\ t_{b_3} = 0{,}13;\ t_{b_4} = 2{,}91;\ t_{b_5} = 0{,}95.$$

За таблицею знаходимо: $t_{0{,}05;\,20-5-1=14} = 2{,}14.$

Таким чином, одержуємо, що тільки $t_{b_4} = 2{,}91 > 2{,}14 = t_{0{,}05;\,14}$. Тоді значущим є коефіцієнт b_4.

За F-критерієм Фішера визначаємо значущість рівняння регресії.

$$F = \frac{R^2(n-p-1)}{(1-R^2)p},$$

де R^2 – коефіцієнт детермінації, що дорівнює 0,517; n = 20; p – кількість факторів, що дорівнює 5.

Виходить, що F=3.

За табл. $F_{0{,}05;\,14} = 2{,}96$. Оскільки $F = 3 > 2{,}96 = F_{0{,}05;\,14}$, то рівняння регресії значуще, але воно визначає залежність тільки від x_4.

Проведемо *покроковий відбір* інформативних факторів.

Складемо таблицю парних коефіцієнтів кореляції за формулою

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_x \cdot S_y}.$$

Таблиця 7.2

	Y	X_1	X_2	X_3	X_4	X_5
Y	1				0,58	
X_1		1	0,85	0,98		
X_2		0,85	1	0,88		0,46
X_3		0,98	0,88	1		
X_4	0,58				1	0,57
X_5			0,46		0,57	1

Примітка. У таблицю занесені тільки значущі коефіцієнти кореляції.

Мультиколінеарність має місце між факторами x_1, x_2, x_3 попарно.

Для усунення мультиколінеарності застосуємо *процедуру покрокового відбору* найбільш інформативних змінних.

Перший крок: найбільший коефіцієнт детермінації має місце для x_4

$$R^2_{y\cdot 4} = (0{,}58)^2 = 0{,}336.$$

Скоректований коефіцієнт детермінації: $\hat{R}^2_{y\cdot 4} = 0{,}299$.

Другий крок: обчислюємо коефіцієнт детермінації для різних пар факторів $(X_4 X_j)(j = 1, 2, 3, 5)$ за формулою

$$R^2 = \frac{b'X'Y' - n(\bar{y})^2}{X'Y - n(\bar{y})^2}. \tag{7.1}$$

Найбільшим виявився $R^2_{Y\cdot X_4 \cdot X_3} = 0{,}483$.

$$\hat{R}^2_{Y\cdot X_4 \cdot X_3} = 0{,}422.$$

Третій крок: за формулою (7.1) обчислюємо коефіцієнт детермінації для різних трійок факторів $(X_4, X_3, X_j)(j = 1, 2, 5)$.

Максимальним коефіцієнтом детермінації виявився

$$R^2_{Y\cdot X_4, X_3, X_5} = 0{,}513.$$

$$\hat{R}^2_{Y \cdot X_4, X_3, X_5} = 0{,}422.$$

Висновок: оскільки скоректований коефіцієнт детермінації не збільшився на третьому кроці, то в регресійній моделі досить обмежитися лише двома раніше відібраними факторами (X_4, X_3) .

Знову розраховане рівняння регресії має вигляд

$$\underset{(0{,}66)}{\hat{y} = 7{,}29} + \underset{(0{,}13)}{3{,}48x_3} + \underset{(1{,}07)}{3{,}48x_4}.$$

Висновок: усі коефіцієнти регресії (b_0, b_3, b_4) значущі, отже, рівняння регресії теж значуще.

Одержувані за допомогою покрокової процедури набори факторів є найбільш оптимальними або близькими до них.

7.2. Лінійні регресійні моделі зі змінною структурою. Фіктивні змінні

На практиці виникає необхідність досліджувати вплив якісних факторів, що мають два або кілька рівнів (градацій). Якісні ознаки можуть істотно впливати на структуру лінійних зв'язків між змінними й приводити до стрибкоподібної зміни параметрів регресійної моделі. У цьому випадку говорять про **дослідження регресійної моделі зі змінною структурою**.

Для опису таких змінних уводять **фіктивні змінні**, тобто використовують *бінарні величини* (0 та 1).

Наприклад, стать

$$z = \begin{cases} 1, \text{ якщо Ч} \\ 0, \text{ якщо Ж} \end{cases}.$$

середня школа, технікум, вища освіта

$$z = \begin{cases} 1, \text{ якщо в/о} \\ \quad 0 \end{cases}.$$

$$z = \begin{cases} 1, \text{ якщо Т} \\ 0 \end{cases}.$$

Приклад 7.2

Необхідно дослідити залежність між результатами вступних і курсових екзаменів з математики.

X – число вирішених задач на вступних іспитах (завдання – 10 задач); Y – число вирішених задач на курсових іспитах (завдання – 7 задач); n – число студентів, n = 12.

Таблиця 7.3

i	x_i	y_i	Стать	i	x_i	y_i	Стать
1	10	6	Ч	7	6	3	Ж
2	6	4	Ж	8	7	4	Ч
3	8	4	Ч	9	9	7	Ч
4	8	5	Ж	10	6	3	Ж
5	6	4	Ж	11	5	2	Ч
6	7	7	Ч	12	7	3	Ж

Рівняння регресійної моделі має вигляд

$$y_i = \beta_0 + \beta_1 x_i + \alpha_1 z_{i_1} + \varepsilon_i .$$

За вихідним даними отримано оцінку цієї моделі:

$$\hat{y} = -1{,}165 + 0{,}743x + 0{,}466z_1,$$

$$\text{де } z_1 = \begin{cases} 1, \text{ якщо Ч} \\ 0, \text{ якщо Ж} \end{cases}.$$

За F-критерієм на рівні $\alpha = 0{,}05$ це рівняння регресії значуще.

Висновок: на курсових іспитах студент вирішує на 0,5 задачі більше, ніж студентка.

За t-критерієм коефіцієнт регресії α_1 при фіктивній змінній z_1 виявився незначущим. Це можливо, якщо:

1) недостатньо вихідних даних;

2) гіпотеза про те, що стать впливає на кількість вирішених задач, неправильна;

3) абітурієнт міг мати освіту (технікум, ПТУ, середня школа).

Тоді регресійна модель має вигляд

$$y_i = \beta_0 + \beta_1 x_1 + \alpha_1 z_{i_1} + \alpha_{21} z_{i_{21}} + \alpha_{22} z_{i_{22}} + \varepsilon_i ,$$

$$z_{21} = \begin{cases} 1, \text{ якщо абітурієнт закінчив сш} \\ \qquad 0, \end{cases}$$

$$z_{22} = \begin{cases} 1, \text{ якщо абітурієнт закінчив сш} \\ \qquad 0. \end{cases}$$

Якщо є дві вибірки, одна з яких невелика, та виникає завдання про їхнє об'єднання, то використовується **критерій Г. Чоу**.

Алгоритм критерію Г. Чоу:

1) За кожною із вибірок будуються лінійні регресійні моделі.

$$y_i = \beta'_0 + \sum_{j=1}^{p} \beta'_j \cdot x_{ij} + \varepsilon'_i, \quad i = 1,2,\ldots n_1$$

$$y_i = \beta''_0 + \sum_{j=1}^{p} \beta''_j \cdot x_{ij} + \varepsilon''_i, \quad i = n+1,\ldots n_1 + n_2 .$$

2) Висувається гіпотеза:

$$\beta' = \beta'', \quad D(\varepsilon') = D(\varepsilon'').$$

3) Обчислюється статистика:

$$F = \frac{\left(\sum_{i=1}^{n} e_i^2 - \sum_{i=1}^{n_1} e_i^2 - \sum_{i=n_1}^{n} e_i^2\right)(n - 2p - 1)}{\left(\sum_{i=1}^{n_1} e_i^2 + \sum_{i=n_1}^{n} e_i^2\right)(p + 1)}, \qquad (7.2)$$

де $\sum_{i=1}^{n} e_i^2, \sum_{i=1}^{n_1} e_i^2, \sum_{i=n_1}^{n} e_i^2$ – залишкові суми квадратів відхилень об'єднаної, першої й другої вибірок, $n = n_1 + n_2$.

4) За таблицею F-критерію знаходимо $F_{\alpha;\, p+1;\, n-2p-2}$, де p – число змінних; n – число спостережень.

5) Якщо $F > F_{\alpha;\, p+1;\, n-2p-2}$, то висунута гіпотеза відкидається.

Приклад 7.3

За даними прикладу 7.2 маємо дві вибірки:

для юнаків: $n_1 = 6; (10;6), (8;4), (7;7), (7;4), (9;7), (5;2);$

для дівчат: $n_2 = 6; (6;4), (8;5), (6;4), (6;3), (6;3), (7;3).$

За формулою (7.2) F = 0,21.

За таблицею знаходимо: $F_{0,05;\, 2;\, 8} = 4{,}46.$

Оскільки F = 0,21<4,46, то висунута гіпотеза приймається. Це означає, що вплив статі студента на результати іспитів несуттєвий.

Контрольні запитання

1. Що означає мультиколінеарність змінних?
2. Який метод застосовується для усунення або зменшення мультиколінеарності?
3. Наведіть алгоритм процедури покрокового відбору.
4. Які особливості лінійних моделей зі змінною структурою?
5. Наведіть алгоритм критерію Г. Чоу.

8. НЕЛІНІЙНІ РЕГРЕСІЙНІ МОДЕЛІ. ЛІНЕАРИЗАЦІЯ МОДЕЛІ. КОЕФІЦІЄНТИ ЧАСТКОВОЇ ЕЛАСТИЧНОСТІ. ЧАСТКОВА КОРЕЛЯЦІЯ

8.1. Нелінійні регресійні моделі. Лінеаризація моделі. Коефіцієнти часткової еластичності

Нехай модель регресії має вигляд

$$y_i = \beta_0 + \beta_1 x_{i1}^2 + \beta_2 \sqrt{x_{i2}} + \varepsilon_i .$$

Це рівняння є *нелінійним за змінними*.

Введемо нові змінні:

$$z_1 = x_1^2 ,\ z_2 = \sqrt{x_2} .$$

Тоді $y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \varepsilon_i$.

Недоліком цієї моделі є труднощі, що виникають при інтерпретації кінцевих результатів.

Розглянемо регресійні моделі, що *нелінійні за параметрами*.

Мультиплікативна (степенева) модель:

$$y_i = \beta_0 \cdot x_{i1}^{\beta_1} \cdot x_{i2}^{\beta_2} \cdot \varepsilon_i .$$

Експонентна модель:

$$y_i = e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i} .$$

Логарифмуючи обидві частини цих рівнянь, одержимо лінійну модель щодо логарифмів змінних.

$$\ln y_i = \ln \beta_0 + \beta_1 \ln x_{i1} + \beta_2 \ln x_{i2} + \ln \varepsilon_i .$$

Для цього рівняння можна застосувати метод найменших квадратів (МНК), але необхідно, щоб вектор збурень ε мав *логарифмічно нормальний розподіл*.

Задача залежності обсягу виробництва від витрат капіталу й праці називається *задачею Кобба–Дугласа*.

$$Y = AK^{\alpha}L^{\beta}\varepsilon, \tag{8.1}$$

де Y – обсяг виробництва; A – коефіцієнт пропорційності; K – витрати капіталу; L – витрати праці; ε – вектор збурень; α, β – **коефіцієнти часткової еластичності** обсягу виробництва за витратами капіталу й праці відповідно.

Коефіцієнтом часткової еластичності $E_{x_i}(y)$ функції $y = f(x_1, x_2, \ldots x_n)$ називається межа відношення відносного часткового збільшення функції до відносного збільшення цієї змінної при $\Delta x_i \to 0$, тобто

$$E_x = \lim_{\Delta x \to 0}\left(\frac{\Delta x y}{y} : \frac{\Delta x}{x}\right) = \frac{x}{y} \cdot y'_x,$$

де y'_x – коефіцієнт еластичності; $E_K = \alpha; E_L = \beta$.

Лінеаризація моделі (8.1) проходить шляхом логарифмування:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \ln \varepsilon.$$

При розширенні виробництва (збільшення витрат капіталу й праці в декілька разів приводить до збільшення обсягу виробництва на таке ж число разів) $\alpha + \beta = 1$.

У цьому випадку $Y = AK^{\alpha}L^{1-\alpha}\varepsilon$:

$$\frac{Y}{L} = A\left(\frac{K}{L}\right)^{\alpha}\varepsilon$$ – *модель розширення виробництва,*

де $\frac{Y}{L}$ – *продуктивність праці*; $\frac{K}{L}$ – *капіталоозброєність виробництва.*

Лінеаризація цієї моделі здійснюється шляхом логарифмування.

Функція Кобба–Дугласа з урахуванням технічного прогресу має вигляд

$$Y = AK^{\alpha}L^{\beta}e^{\theta \cdot t}\varepsilon,$$

де θ – параметр темпу приросту обсягу виробництва завдяки технічному прогресу; t – час.

8.2. Часткова кореляція

Частковим коефіцієнтом кореляції між змінними X_i, X_j називається вираз

$$r_{ij.12\ldots p} = \frac{-q_{ij}}{\sqrt{q_{ii}q_{jj}}},$$

де q_{ij} – алгебричне доповнення елементів r_{ij} матриці парних коефіцієнтів кореляції.

$$q_p = \begin{pmatrix} 1 & r_{12} & \ldots & r_{1p} \\ r_{21} & 1 & \ldots & r_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ r_{p1} & r_{p2} & \ldots & 1 \end{pmatrix},$$

де $r_{ij} = \dfrac{\sum\limits_{i=1}^{n}\left(x_i - \bar{x}_i\right)\left(x_j - \bar{x}_j\right)}{n \cdot S_{x_i} \cdot S_{x_j}}$.

Для трьох змінних

$$r_{ij.k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{\left(1 - r_{ik}^2\right)\left(1 - r_{jk}^2\right)}}.$$

Значущість часткового коефіцієнта кореляції оцінюється так само, як і для парного коефіцієнта кореляції, припускаючи при цьому $n' = n - p + 2$.

Приклад 8.1

Дослідити залежність між продуктивністю праці (X_1), віком (X_2) і виробничим стажем (X_3) працівника.

Обчислені парні коефіцієнти кореляції виявилися значущими:

$$r_{12} = 0{,}2; \quad r_{13} = 0{,}41; \quad r_{23} = 0{,}82.$$

$$q_3 = \begin{pmatrix} 1 & 0,2 & 0,41 \\ 0,2 & 1 & 0,82 \\ 0,41 & 0,82 & 1 \end{pmatrix}.$$

Часткові коефіцієнти кореляції дорівнюють:

$$r_{12.3} = \frac{0,2 - 0,41 \cdot 0,82}{\sqrt{\left(1 - 0,41^2\right)\left(1 - 0,82^2\right)}} = -0,26;$$

$$r_{13.2} = 0,44; \quad r_{23.1} = 0,83.$$

Виключення якого-небудь фактора в економетриці називається *елімінуванням*.

Контрольні запитання

1. Які існують види нелінійних регресійних моделей?
2. Дайте визначення поняття виробничої функції.
3. У чому суть коефіцієнта часткової еластичності?
4. Перелічіть види виробничої функції.
5. Назвіть характеристики виробничої функції.
6. У чому суть часткового коефіцієнта кореляції?

9. ЧАСОВІ РЯДИ

9.1. Основні поняття

Часові ряди – це сукупність значень будь-якої ознаки за кілька послідовних моментів або періодів часу.

Окремі значення цієї ознаки називаються *рівнями ряду*.

Рівні часового ряду визначаються безліччю факторів, які можна розбити на кілька груп:

1. Фактори, що формують тенденцію ряду або його **тренд** (T).
2. Циклічні коливання (S).
3. Випадкові фактори (E).

Найбільше часто використовують:

1) **адитивну модель**

$$Y = T + S + E;$$

2) **мультиплікативну модель**

$$Y = T \cdot S \cdot E.$$

9.2. Автокореляція рівнів часового ряду

У загальному випадку наступні рівні часового ряду залежать від попередніх.

Кореляційна залежність між послідовними рівнями ряду називається *автокореляцією рівнів ряду*.

Кількісно вона оцінюється за допомогою лінійного парного коефіцієнта кореляції між рівнями, зрушеними один відносно одного. Крок зрушення називається *лагом*.

Приклад 9.1

Є дані про середні витрати на кінцеве споживання за вісім років (табл. 9.1).

Таблиця 9.1

t	1	2	3	4	5	6	7	8
y_t	7	8	8	10	11	12	14	16
y_{t-1}	–	7	8	8	10	11	12	14

Коефіцієнт кореляції обчислюється за формулою

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}.$$

Тоді ця формула прийме вигляд

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum(y_t - \bar{y}_1)^2 \sum(y_{t-1} - \bar{y}_2)^2}}. \quad \bar{y}_1 = \sum_{t=2}^{n} y_t /(n-1); \quad \bar{y}_2 = \sum_{t=2}^{n} y_{t-1} /(n-1).$$

Величина r_1 називається коефіцієнтом автокореляції рівнів ряду першого порядку (лаг = 1)

$$\bar{y}_1 = \frac{8+8+\ldots+16}{7} = 11{,}29;$$

$$\bar{y}_2 = \frac{7+8+\ldots+14}{7} = 10;$$

$$r_1 = 0{,}975.$$

Для коефіцієнта автокореляції другого порядку використовуємо такі зрушені ряди (табл. 9.2):

Таблиця 9.2

t	1	2	3	4	5	6	7	8
y_t	7	8	8	10	11	12	14	16
y_{t-2}	–	–	7	8	8	10	11	12

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}}. \quad \bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}; \quad \bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}.$$

Обчислений $r_2 = 0{,}973$.

Аналогічно обчислюємо коефіцієнти автокореляції 3-го, 4-го та інших порядків.

Для одержання статистично достовірних результатів вважають, що максимальний лаг не повинен перевищувати величину (n : 4).

Послідовність коефіцієнтів автокореляції називається **автокореляційною функцією**, а її графік називається **корелограмою**.

Аналіз автокореляційної функції дозволяє зробити деякі висновки:

1. Якщо найбільшим виявився r_1, то часовий ряд містить тільки тенденцію.

2. Якщо найбільшим виявився r_τ, то ряд містить циклічне коливання з періодом τ.

3. Якщо жоден із коефіцієнтів автокореляції не є значимим, то ряд:

- або містить тільки випадкову складову;
- або містить сильну нелінійну тенденцію.

9.3. Моделювання часового ряду

Моделювання часового ряду зводиться до розрахунку значень *T*, *S*, *E* для кожного рівня ряду.

Цей процес включає наступні кроки:

1) вирівнювання часового ряду методом ковзної середньої (табл. 9.3).

Таблиця 9.3

t	1	2	3	4	5	6	7	8
y_t	213	171	291	309	317	362	351	361
$\tilde{y}_t$	–	225	241	305,7	329,3	336,3	358	–

$$\tilde{y}_2 = \frac{y_1 + y_2 + y_3}{3} = (213 + 171 + 291)/3 = 225;\ \ \tilde{y}_3 = \frac{y_2 + y_3 + y_4}{3} = 241;$$

2) розрахунок значень циклічного компонента *S*;

3) усунення *S* з вихідних рівнів часового ряду;

4) аналітичне вирівнювання рівнів ряду (одержання рівняння тренда) і розрахунок значень *T* з використанням отриманого рівняння тренда;

5) розрахунок отриманих за моделлю значень *T* + *S* або *T* · *S*;

6) розрахунок помилок.

Приклад 9.2

Є дані про споживання електроенергії в районі за чотири роки (табл. 9.4).

Таблиця 9.4

Номер кварталу t	Споживання електроенергії, y_t	Σ за чотири квартали	Ковзне середнє за чотири квартали	Центроване ковзне середнє	Циклічний компонент S_t
1	2	3	4	5	6
1	6	–	–	–	–
2	4,4	24,4	6,1	–	–
3	5	25,6	6,4	6,25	–1,25
4	9	26	6,5	6,45	2,55
5	7,2	27	6,75	6,625	0,575
6	4,8	28	7	6,875	–2,075
…	…	…	…	…	…
16	10,8	–	–	–	–

З аналізу вихідних рівнів ряду видно, що амплітуда коливань приблизно постійна, тому необхідно скористатися *адитивною моделлю*:

$$Y = T + S + E.$$

КРОК 1: Вирівняємо вихідні рівні методом ковзної середньої. Для цього:

а) підсумуємо рівні ряду послідовно за кожні чотири квартали зі зрушенням на 1 квартал (графа 3.4: 3 – номер стовпця, 4 – номер рядка);

б) розділивши отримані суми на 4, знайдемо значення ковзної середньої (графа 4.4). Ці значення вже не містять циклічного компоненту;

в) приведемо ці значення у відповідність до фактичних моментів часу, для чого знайдемо середнє значення із двох сусідніх значень – центровані значення ковзної середньої (графа 5.4).

КРОК 2: Знайдемо оцінки циклічного компонента S_t як різниці між y_t і центрованими значеннями ковзної середньої (графа 6.4).

У моделях із циклічним компонентом звичайно передбачається, що циклічні впливи за період взаємопогашаються. В адитивній моделі це означає, що сума значень $\bar{S}_i$ за всіма кварталами повинна дорівнювати нулю.

$$\sum \bar{S}_i = 0{,}6 - 1{,}958 - 1{,}275 + 2{,}708 = 0{,}075.$$

Коригувальне виправлення визначається за формулою

$$K = \frac{0{,}075}{4} = 0{,}01875.$$

Скоректовані значення S_i: $S_i = \bar{S}_i - K$.

Отримані значення S_i представлені в табл. 9.5

Таблиця 9.5

Показник	Рік	Номер кварталу (i)			
		I	II	III	IV
	1	–	–	–1,25	2,55
	2	0,575	–2,075	–1,1	2,7
	3	0,55	–2,025	–1,475	2,875
	4	0,675	–1,775	–	–
Разом за i-й квартал (за чотири роки)		1,8	–5,875	–3,825	8,125
Середня оцінка $\bar{S}_i$		0,6	–1,958	–1,275	2,708
Скоректована оцінка S_i		0,581	1,977	–1,294	2,69

КРОК 3: Елімінуємо (видалимо) вплив циклічного компонента, віднімаючи його значення з рівнів часового ряду y_t:

$$(T+E)_i = y_t - S_i .$$

Таблиця 9.6

t	y_t	S	$T + E$	T	$T + S$	E	E^2
1	2	3	4	5	6	7	8
1	6	0,581	5,419	5,902	6,483	–0,483	0,2333
2	4,4	–1,977	6,337	6,088	4,111	0,289	0,0835
3	5	–1,294	6,294	6,275	4,981	0,019	0,0004
4	9	2,69	6,31	6,461	9,151	–0,151	0,0228
5	7,2	0,581	6,619	6,648	7,229	–0,29	0,0008
6	4,8	–1,977	6,777	6,834	4,857	–0,057	0,0032
…	…	…	…	…	…	…	…
16	10,8	2,69	8,11	8,698	11,388	–0,588	0,3457

КРОК 4: Визначимо компоненту *T*. Для цього проведемо аналітичне вирівнювання часового ряду *T* + *E* (графа 4.6: 4 – номер стовпця, 6 – номер рядка) за допомогою лінійного тренду.

За формулами розділу 2 для парної лінійної регресії:

$$T = 5{,}715+0{,}186t.$$

У цій формулі замість змінної *x* був підставлений час *t*.

Із цього рівняння знайдемо значення *t* для кожного моменту часу (графа 5.6).

КРОК 5: Обчислимо значення абсолютних відхилень (графа 7.6) та їхні квадрати (графа 8.6).

За даними графи 8.6 знайдемо коефіцієнт детермінації:

$$Q = \sum_t \left(y_t - \bar{y}_t\right)^2;$$

$$Q_e = \sum_t E^2 = 1{,}1.$$

Тоді $R^2 = \dfrac{1 - 1{,}1}{71{,}59} = 0{,}985.$

Це означає, що адитивна модель пояснює 98,5 % загальної варіації рівнів часового ряду споживання електроенергії.

Приклад 9.3

Знайти прогноз на споживання електроенергії в першому півріччі наступного року.

T_{17} = 5,715 + 0,186·17 = 8,877;

T_{18} = 5,715 + 0,186·18 = 9,063.

S_{17} = S_{I} = 0,581;

S_{18} = S_{II} = –1,977.

Тоді $F = T_{17} + T_{18} + S_{17} + S_{18}$ = 16,544 млн кВт/г.

Аналогічним чином будують *мультиплікативну модель*.

Контрольні запитання

1. У чому особливості побудови моделей динаміки?
2. Які складові можуть бути виділені в часовому ряді?
3. Дайте визначення автокореляції.
4. У чому суть автокореляційної функції?
5. Які висновки випливають із аналізу автокореляційної функції?
6. Наведіть алгоритм моделювання часового ряду.

10. РОЗРАХУНКОВО-ГРАФІЧНІ ЗАВДАННЯ

10.1. Парний регресійний аналіз

Вказівки до виконання: є наступні дані про значення ціни на товар (грн) і обсяг пропозиції (попиту) – тис. грн. Необхідно побудувати лінійні однофакторні економетричні моделі та привести їхні графіки, розрахувати їхні характеристики, зробити висновки щодо їхньої адекватності, розрахувати прогнозні значення, якщо ціна складе 7,2 грн.

Варіант 1

X	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	13,7	18,0	6,2	15,5	24,1	24,8	25,0	13,0	8,1	6,7

Варіант 2

X	5,5	7,1	2,6	5,4	10,0	11,6	12,4	2,9	2,4	1,6
Y	17,7	14,0	4,2	12,5	29,1	25,8	22,0	9,0	4,1	3,7

Варіант 3

X	13,7	18,0	6,2	15,5	24,1	24,8	25,0	13,0	8,1	6,7
Y	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6

Варіант 4

X	5,1	7,4	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	16,7	13,0	7,2	14,5	20,1	21,8	23,0	11,0	7,1	4,7

Варіант 5

X	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	13,7	18,0	6,2	15,5	24,1	24,8	25,0	13,0	8,1	6,7

Варіант 6

X	15,4	17,6	12,3	15,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	13,7	18,0	16,2	19,5	14,1	14,8	15,0	9,0	5,1	3,7

Варіант 7

X	5,1	7,2	2,4	5,1	11,2	12,2	10,1	4,8	2,3	1,5
Y	18,7	28,0	16,2	25,5	34,1	44,2	35,0	23,0	8,1	10,7

Варіант 8

X	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	11,7	14,0	5,2	13,5	20,1	21,8	24,0	12,0	7,1	6,7

Варіант 9

X	5,3	7,4	2,4	5,6	11,1	12,2	10,3	4,8	2,3	1,5
Y	33,1	38,0	16,2	40,5	54,1	64,8	55,0	53,0	15,1	10,7

Варіант 10

X	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	3,7	8,0	1,2	3,5	9,1	8,8	9,2	3,0	1,1	0,7

Варіант 11

X	3,4	5,6	1,3	3,9	8,0	10,6	9,4	3,9	1,4	1,0
Y	13,7	18,0	6,2	15,5	24,1	24,8	25,0	13,0	8,1	6,7

Варіант 12

X	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	1,7	1,8	1,6	2,5	7,1	6,8	5,0	2,0	1,7	0,7

Варіант 13

X	3,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	13,7	28,0	9,7	19,5	44,1	54,8	55,0	43,0	18,1	16,7

Варіант 14

X	2,4	7,8	2,1	5,8	11,2	12,5	10,3	4,8	2,3	1,5
Y	13,7	38,0	16,2	35,5	54,1	64,8	55,0	23,0	15,1	10,7

Варіант 15

X	0,4	0,6	0,3	0,9	1,0	1,6	1,4	0,6	0,3	0,6
Y	13,7	18,0	6,2	15,5	24,1	24,8	25,0	13,0	8,1	16,7

Варіант 16

X	1,4	2,6	2,3	5,1	6,0	5,6	10,4	4,9	2,4	1,6
Y	12,7	18,0	16,2	25,5	24,1	24,8	55,0	33,0	18,1	16,7

Варіант 17

X	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	3,7	8,0	3,2	2,5	4,9	4,8	5,1	2,2	1,1	0,7

Варіант 18

X	5,1	7,2	2,0	5,1	10,0	11,6	9,4	6,9	3,4	2,6
Y	15,7	17,0	5,2	17,5	22,1	25,8	23,0	17,0	9,1	7,7

Варіант 19

X	4,4	6,6	3,3	4,9	11,0	11,6	9,4	3,9	5,4	2,6
Y	18,7	28,0	16,2	26,5	44,1	44,8	35,0	16,0	12,1	7,7

Варіант 20

X	5,3	7,5	2,1	5,8	10,0	11,3	8,4	1,9	7,4	1,1
Y	43,7	58,0	16,2	55,5	64,1	74,8	65,0	13,0	38,1	6,7

Варіант 21

X	5,1	7,4	2,2	5,7	11,1	12,3	10,2	4,9	2,4	1,6
Y	8,7	14,0	7,2	13,5	14,1	14,8	15,0	6,0	4,1	4,7

Варіант 22

X	5,1	7,3	2,2	5,8	12,0	10,6	11,4	3,9	1,4	2,6
Y	13,5	18,2	6,3	15,1	24,2	24,4	25,2	13,1	8,5	6,3

Варіант 23

X	5,2	7,3	2,4	5,6	11,2	12,2	10,2	4,3	2,3	1,5
Y	12,7	16,0	5,8	13,5	21,1	22,8	23,0	15,1	7,1	5,7

Варіант 24

X	5,8	5,6	5,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	13,0	18,4	16,2	19,5	34,1	40,0	35,0	19,2	10,4	9,7

Варіант 25

X	5,7	7,1	3,2	5,7	9,1	12,3	10,2	4,9	2,4	1,6
Y	8,1	13,6	7,7	11,5	22,1	24,8	19,6	6,0	4,7	4,9

10.2. Множинний регресійний аналіз

Вказівки до виконання: побудувати лінійну двофакторну модель $Y_i = f(X_1, X_2)$, оцінити параметри моделі й перевірити їхню статистичну значущість за допомогою критерію Ст’юдента, перевірити значущість моделі за критерієм Фішера, знайти коефіцієнт детермінації й множинної кореляції, розрахувати коефіцієнти парної кореляції між факторними ознаками. Зробити висновки щодо адекватності моделі.

Таблиця 10.1

i	X_1	X_2	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
1	2,2	5,4	10	18	38	47	20,9	52,3	19,4	30	46,7	34,3
2	3,1	3,2	14	12	16	26	14,6	33	15,4	17	30,3	23,7
3	5,4	5,6	15	28	34	36	28,1	42,4	36,3	28	39,7	30,8
4	6,1	7,8	12,8	40	52	51	36	57,5	44	38	51,3	36,9
5	7,8	3,5	25	28	10	10	26,5	16,7	40,3	11	20,5	15,4
6	5,6	6,8	12,1	35	44	44	30,8	50,3	41,5	31	46,8	34,7
7	4,2	4,8	15	23	29	34	22,4	38,9	26,6	23	38,5	25,6
8	4,4	5,1	14,7	24	31	36	22,1	42,7	30,5	26	41,1	27,4
9	2,2	3,2	13,1	9	19	30	12,8	34,6	12	21	36,5	22,9
10	3,1	4,6	17	19	29	37	19,5	44,3	22,2	28	38,9	28,1

Таблиця 10.2

i	X_1	X_2	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}
1	3,2	5,8	9	11,2	37	22	17,2	17,3	33,5	10,2	30	4
2	5,3	6,9	15	12,7	37	23	10,1	26,7	42,2	13,3	43	8
3	6,7	8,1	20	12,4	40	26	9,6	31,5	52,4	16,4	51	8
4	2,2	3,2	8	14,9	78	16	13,9	4,9	13,9	9	18	6
5	8,8	9,6	25	17,3	41	26	1,7	42	65,3	20,2	62	9
6	4,3	5,6	13	13,6	34	20	9,8	18	33,2	12,9	33	7
7	6,9	9,3	20	11,2	44	32	9,5	38	61,7	15,9	55	6
8	2,1	3,5	9	14,7	30	14	14,8	6	16,3	10	19	5
9	8,9	12	24	8,1	50	38	6,1	53	81,2	18,6	72	5
10	10,1	15	26	3,7	60	44	3,8	65	102,8	20	88	4

Таблиця 10.3

i	X_1	X_2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	Y_{25}
1	2,3	3,5	4,8	9	15	21	10
2	7,2	8,3	26	16,5	43,1	52	43
3	2,5	4,3	6,1	7,1	16,1	27	15
4	3,7	6,9	12,5	9,1	26,1	39	31
5	6,1	7,1	21,1	15,5	36,4	43	35
6	2,5	5,7	7,2	8,9	19	33	21
7	5,3	7,2	17,4	13,5	33	42	36
8	2,1	5,6	6	5,5	17	32	19
9	4,5	7,1	15,7	11,6	30	42	32
10	6,2	8,6	21,2	13,9	40	50	43

10.3. Лінійна регресійна модель зі змінною структурою. Фіктивні змінні

Вказівки до виконання: є наступні дані про вагу Y (у фунтах) та вік X (у тижнях) 11 індичок, вирощених у США й Канаді. Необхідно для свого варіанта вихідних даних:

1) знайти рівняння парної регресії Y за X та оцінити його значущість;

2) ввівши відповідні фіктивні змінні, знайти загальне рівняння множинної регресії Y за всіма пояснюючими змінними (включаючи фіктивні);

3) оцінити значущість загального рівняння множинної регресії за F-критерієм і значущість його коефіцієнтів за t-критерієм.

Коефіцієнт варіанта завдання визначається за формулою

$$k = \frac{100 + N}{100},$$

де N – номер прізвища студента в журналі групи.

Таблиця 10.4

i	x_i	y_i	Країна походження
1	28	12,3*k	Канада
2	20	8,9*k	Канада
3	32	15,1*k	Канада
4	22	10,4*k	Канада
5	29	13,1*k	Канада
6	27	12,4*k	США
7	28	13,2*k	США
8	26	11,8*k	США
9	21	11,5*k	США
10	27	14,2*k	Канада
11	29	15,4*k	США

10.4. Нелінійні регресійні моделі. Функція Кобба–Дугласа

Вказівки до виконання: є класична модель виробничої функції Кобба–Дугласа, що описує залежність між випуском продукції (Y), обсягом капіталу (X_1) і обсягом трудових ресурсів (X_2).

$$Y = \beta_0 X_1^{\beta 1} X_2^{\beta 2}.$$

Необхідно для свого варіанта вихідних даних оцінити параметри моделі; розрахувати коефіцієнт множинної кореляції.

Таблиця 14.5

N	X_1, *K* (млн грн)	X_2, *L* (тис. чол.)	*Y*, ВП (млн грн)
1	1,3*k*	2,2*k*	2,15*k*
2	2,1*k*	2,6*k*	4,41*k*
3	2,6*k*	3,3*k*	5,54*k*
4	3,1*k*	3,6*k*	6,69*k*
5	4,2*k*	4,6*k*	7,48*k*
6	4,6*k*	5,2*k*	9,56*k*
7	5,3*k*	5,8*k*	10,62*k*
8	5,7*k*	6,2*k*	11,94*k*
9	6,8*k*	7,5*k*	12,02*k*
10	8,4*k*	10,8*k*	18,51*k*

10.5. Часові ряди

Вказівки до виконання: для значень представленого ряду необхідно побудувати модель декомпозиції часового ряду, виділити такі складові, як тренд, циклічна, сезонна та випадкова компоненти. Побудувати прогноз для 2003 р. за кварталами з урахуванням тренду, циклічної та випадкової складових.

Таблиця 10.6

		Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
1990	1	23,4	73,526	40,69	101,5	85,06	176,8	42,163	56,7	118,42	27,915
	2	33,7	114,67	68,76	92,14	126,2	145,7	46,703	80,4	114,92	26,85
	3	44,1	134,91	91,76	89,74	152,9	133,2	43,59	91	126,36	25,612
	4	20,4	61,359	45,03	79,43	66,91	120	30,41	39,5	62,481	21,517
1991	1	34,4	80,079	67,69	191,5	97,02	265,5	44,45	51,3	129,24	46,061
	2	44,2	123,19	115,9	154,8	144,6	220,5	53,20	81,5	115,03	32,727
	3	61,5	146,81	152,7	144	168	194,3	54,393	103	113,47	28,806
	4	33,5	64,89	73,84	119,3	79,35	166,9	28,66	46,6	76,189	32,263
1992	1	44,9	88,115	110,4	275	103,6	363,6	48,905	66	108,91	60,59
	2	64,4	136,37	184,2	216,4	161	284,2	50,719	108	114,26	45,861
	3	82,1	162,15	240,1	189,5	185	253,6	57,38	129	133,86	48,957
	4	38,4	74,031	114,2	155,2	83,34	205,4	30,32	59,2	67,08	42,676
1993	1	60,5	100,06	168,8	348,3	110,8	466,8	62,99	84,3	129,19	80,661
	2	85,5	152,32	279,6	269,5	178,6	358,1	58,177	138	125	69,214
	3	108	181,3	359,6	229,1	206,8	316,5	62,695	179	127,03	61,839
	4	50,8	80,12	171	186,6	98,59	256,6	35,469	89	71,498	47,219
1994	1	76	109,93	248,7	404,8	128,7	566,1	66,88	124	126,79	103,18
	2	116	171,01	405,3	304,7	195,4	440,7	65,929	192	135,31	86,968
	3	143	203,29	514,9	254,5	230,9	377,3	67,836	243	118,93	76,189
	4	65,8	89,795	240,6	201	104,2	303	37,083	116	75,95	61,543
1995	1	93,1	125,09	347,9	434,3	141,5	681,4	66,794	164	134,23	131,23
	2	147	193,76	562	321,6	209,9	523,2	61,867	272	131,98	105,43
	3	177	230,67	708,2	264,4	253,6	443,4	70,592	339	130,89	92,286
	4	84,6	103,22	327,3	203,5	111,5	362,9	38,671	160	81,023	73
1996	1	114	141,76	468	433,1	149,8	790,9	63,40	225	116,18	163,85
	2	177	217,69	754,8	314,6	231,3	597,5	62,487	367	132,1	130,09
	3	223	262,63	943	252,7	271,6	513,2	67,065	456	113,18	111,96
	4	102	118,72	433,4	189,2	126,4	412,8	36,351	209	66,336	86,87
1997	1	147	161,66	615,8	395,2	166,1	903,4	66,805	303	127,77	198,52
	2	218	250,64	983,2	275,1	248,4	686,5	64,213	481	123,74	154,23
	3	273	301,54	1221	217,1	304,1	576,6	68,974	609	125,87	134,08
	4	120	132,98	556,9	157,5	137,4	463	43,233	278	63,535	106,13
1998	1	170	185,78	787,5	307,5	180,4	1024	64,638	392	123,7	233,14
	2	264	283,73	1253	202,2	279,8	772	65,279	629	129,27	179,32
	3	332	342,69	1548	148,5	326,1	651,8	59,24	787	127,18	154,61
	4	153	152,68	702,5	99,44	144	522,3	38,887	360	81,884	127,58

Таблиця 10.7

		Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}
1993	1	26	74	40,68	101,5	85	177	42,16	56,69	118,4	27,9
	2	31	115	68,76	92,14	126	146	36,7	80,36	114,9	26,8
	3	45	125	91,76	89,74	153	133	43,59	91	126,4	25,6
	4	22	71	45,03	79,43	67	120	30,41	39,51	62,48	21,5
1994	1	31	80	67,68	191,5	97	265	44,46	51,27	129,2	46,1
	2	48	123	115,9	154,8	145	221	53,21	81,55	115	32,7
	3	63	147	152,7	144	168	194	44,39	102,9	113,5	28,8
	4	35	55	73,83	119,3	79	167	28,66	46,64	76,19	32,3
1995	1	48	88	110,4	275	104	364	48,91	66,05	108,9	60,6
	2	62	126	184,2	216,4	161	284	54,72	107,9	114,3	45,9
	3	88	162	240,1	189,5	185	254	57,39	129	133,9	49
	4	32	74	114,2	155,2	83	205	30,32	59,21	67,08	42,7
1996	1	64	100	168,8	348,3	111	467	62,99	84,26	129,2	80,7
	2	88	152	279,5	269,5	179	358	58,18	138,2	125	69,2
	3	102	181	359,6	229,1	207	317	68,7	179	127	61,8
	4	56	80	170,9	186,6	99	257	35,47	88,96	71,5	47,2
1997	1	76	116	248,6	404,8	129	566	66,88	124,2	126,8	103
	2	117	171	405,3	304,7	195	441	65,93	192,2	135,3	87
	3	143	203	514,9	254,5	231	377	67,84	243,3	118,9	76,2
	4	66	91	240,6	201	104	303	37,08	116,3	75,95	61,5
1998	1	93	125	347,8	434,3	141	681	66,79	163,8	134,2	131
	2	147	194	562,0	321,6	210	523	61,87	272	132	105
	3	177	233	708,1	264,4	254	443	75,59	340,7	130,9	92,3
	4	85	103	327,3	203,5	112	363	38,67	159,7	81,02	73
1999	1	114	142	468,0	433,1	150	791	63,4	225	116,2	164
	2	177	218	754,7	314,6	231	598	62,49	356,8	132,1	130
	3	213	263	942,9	252,7	272	513	67,06	456,4	113,2	112
	4	112	119	433,4	189,2	126	413	36,35	209,1	66,34	86,9
2000	1	167	164	615,8	395,2	166	903	66,8	302,7	127,8	199
	2	218	257	983,2	275,1	248	686	64,21	480,7	123,7	154
	3	273	312	1221,	217,1	304	577	68,97	609,5	125,9	134
	4	120	133	556,9	157,5	137	463	43,23	277,8	63,53	106
2001	1	180	186	787,4	307,5	180	1024	64,64	391,8	123,7	233
	2	264	284	1252,	202,2	280	772	65,28	628,5	129,3	179
	3	332	343	1548,	148,5	326	652	59,24	786,7	127,2	155
	4	153	143	702,4	99,44	144	522	38,89	359,6	81,88	128

Таблиця 10.8

		Y_{21}	Y_{22}	Y_{23}	Y_{24}	Y_{25}
1992	1	23	74	40,686	101,5	85
	2	34	115	68,765	92,14	126
	3	44	135	91,761	89,74	153
	4	20	61	45,033	79,43	67
1993	1	34	80	67,687	191,5	97
	2	44	123	115,93	154,8	145
	3	62	147	152,71	144	168
	4	34	65	73,837	119,3	79
1994	1	45	88	110,44	275	104
	2	64	136	184,25	216,4	161
	3	82	162	240,14	189,5	185
	4	38	74	114,23	155,2	83
1995	1	60	100	168,81	348,3	111
	2	85	152	279,57	269,5	179
	3	108	181	359,61	229,1	207
	4	51	80	170,95	186,6	99
1996	1	76	110	248,67	404,8	129
	2	116	171	405,33	304,7	195
	3	143	203	514,94	254,5	231
	4	66	90	240,64	201	104
1997	1	93	125	347,86	434,3	141

ДОДАТОК 1

Розподіл Фішера

У таблицях Д 1.1, Д 1.2 наведено критичні значення F_{ν_1,ν_2} залежно від числа ступенів вільності ν_1 й ν_2 для рівня значущості $\alpha = 0{,}05$.

Таблиця Д 1.1

ν_2	ν_1									
	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	3,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

Таблиця Д 1.2

v_2	v_1								
	12	15	20	24	30	40	60	120	∞
1	244	246	248	249	250	251	252	253	254
2	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
30	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39

ДОДАТОК 2

Критичні точки розподілу Ст'юдента

У таблиці Д 2.1 наведено значення *t*-критерію Ст'юдента залежно від числа ступенів вільності v і рівня значущості α.

Таблиця Д 2.1

Число ступенів вільності v	Рівень значущості α					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31.6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1.94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,80	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,70	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4.32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,95
18	1,73	2,10	2.55	2,88	3.61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
25	1.71	2,06	2,49	2,79	3,45	3,72
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29

ДОДАТОК 3

Критичні точки розподілу χ^2 Пірсона

У таблиці Д 3.1 наведено значення $\chi^2_{\alpha,\nu}$ залежно від числа ступенів вільності ν і рівня значущості α.

Таблиця Д 3.1

ν	α					
	0,20	0,10	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,266
4	5,989	7,779	9,488	11,668	13,277	18,467
5	7,289	9,236	11,070	13,388	15,086	20,515
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19,679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,725	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528
14	18,151	21,064	23,685	26,783	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	33,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,687	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45,315
22	27,301	30,813	33,924	37,659	40,289	48,268
24	29,553	33,196	36,415	40,270	42,980	51,179
26	31,795	35,563	38,885	42,856	45,642	54,052
28	34,027	37,916	41,337	45,419	48,278	56,893
30	36,250	40,256	43,773	47,962	50,892	59,703

СПИСОК ЛІТЕРАТУРИ

1. Клебанова Т. С. Эконометрия : учебное пособие / Т. С. Клебанова, Н. А. Дубровина, Е. В. Раевнева. – Харьков : ИНЖЭК, 2005. – 160 с.
2. Кремер Н. Ш. Эконометрика : учебник для вузов / Н. Ш. Кремер, Б. А. Путко; под ред. проф. Н. Ш. Кремера. – М. : ЮНИТИ-ДАНА, 2005. – 311 с.
3. Доугерти К. Ведение в эконометрику: учебник / К. Догерти. – М. : ИНФРА-М, 2004. – 432 с.
4. Елисеева И. И. Эконометрика: учебник для вузов / И. И. Елисеева, С. В. Курышева, Т. В. Костеева; под ред. чл.-корр. РАН И. И. Елисеевой. – М. : Финансы и статистика, 2002. – 342 с.
5. Елисеева И. И. Практикум по эконометрике / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко; под ред. чл.-корр. РАН И. И. Елисеевой. – М. : Финансы и статистика, 2002. – 191 с.
6. Наконечний С. І. Економетрія : підручник / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – К. : КНЕУ, 2005. – 520 с.
7. Лугінін О. Є. Економетрія: навчальний посібник / О. Є. Лугінін, С. В. Білоусова, О. М. Білоусов. – К. : Центр навчальної літератури, 2005. – 252 с.

Навчальне видання

БІЛОЦЕРКІВСЬКИЙ Олександр Борисович

ШИРЯЄВА Наталя Володимирівна

ЕКОНОМЕТРІЯ

Навчально-методичний посібник

для студентів спеціальності 7.050206

"Менеджмент зовнішньоекономічної діяльності"

Роботу до видання рекомендував *В. А. Міщенко*

Редактор *Н. В. Ковшарь*

План 2008 р., поз. 125/184-08
Підписано до друку 1.12.08. Формат $60 \times 84\ {}^{1}\!/_{16}$. Папір офсетний.
Друк - ризографія. Гарнітура Таймс. Ум. друк. арк. 3,8. Обл.-вид. арк. 4,0.
Наклад 150 прим. Зам. № 336. Ціна договірна.

Видавничий центр НТУ "ХПІ".
Свідоцтво про державну реєстрацію ДК № 116 від 10.07.2000 р.
61002, Харків 2, вул. Фрунзе, 21

Друкарня НТУ "ХПІ". 61002, Харків, вул. Фрунзе, 21